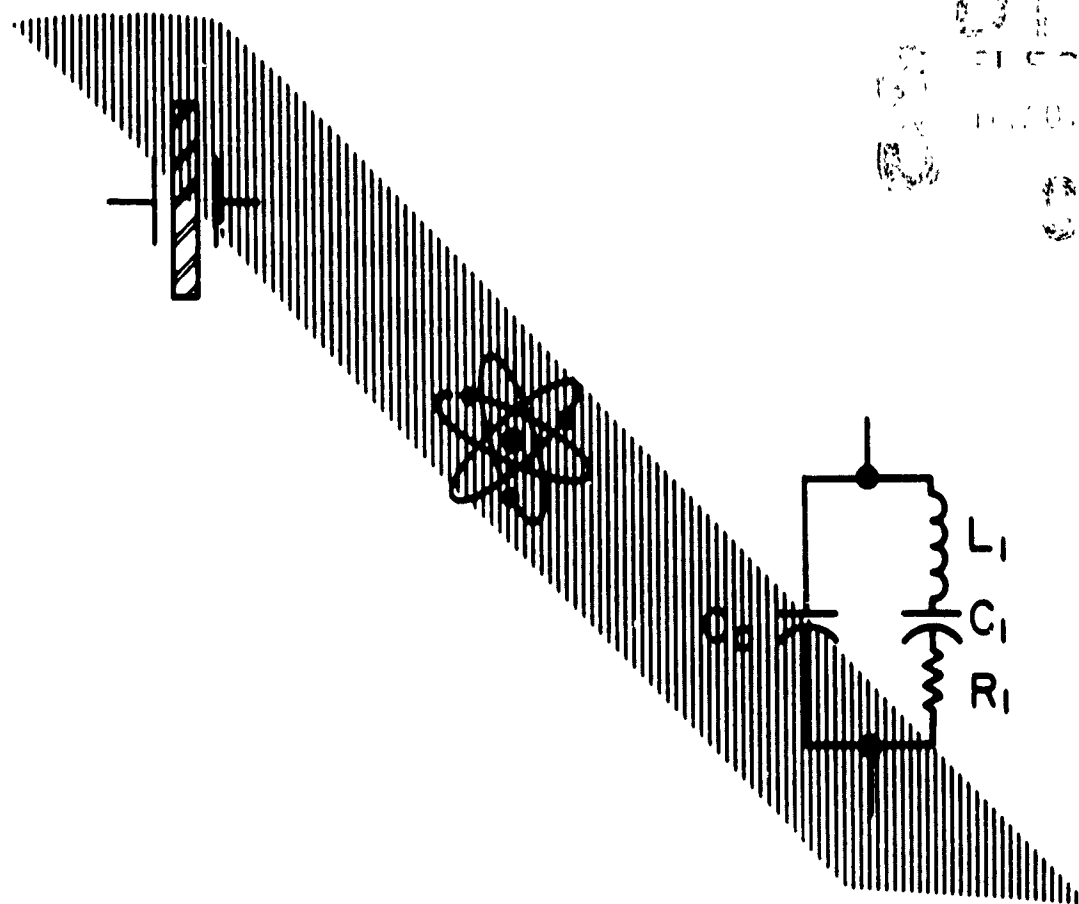


AD-A235 435



2

PROCEEDINGS
OF THE
40TH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL
1986



28 - 30 MAY 1986

86CH2330-9

58-60781

NOTE TO DTIC for CITATION

→ Suggested Descriptors:

accelerometers
aging
Allan variance
atomic clocks
atomic frequency standards
atomic resonators
cesium
cesium standards
clocks
crystals
etching
filters
frequency control
frequency standards
hydrogen masers
noise(electrical and electromagnetic)
oscillators
phase noise
quartz
quartz crystals
quartz oscillators
quartz resonators
radiation effects
resonators
rubidium
rubidium standards
sensors
stability
surface acoustic wave devices
time
timekeeping
transducers
vibration effects

(Note: Headings in Table of Contents are good
to use for Abstract information.)

**PROCEEDINGS
OF THE
40th ANNUAL
FREQUENCY SYMPOSIUM
1986**

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By <i>Res Form 50</i>	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
<i>A1</i>	



Co-Sponsored by the



U.S. ARMY
ELECTRONICS TECHNOLOGY AND
DEVICES LABORATORY

and



THE INSTITUTE OF ELECTRICAL AND
ELECTRONICS ENGINEERS, INC.
ULTRASONICS, FERROELECTRICS AND
FREQUENCY CONTROL SOCIETY

IEEE Catalog No. 86CH2330-9

Library of Congress No. 58-60781

Dunfey City Line Hotel

Philadelphia, Pennsylvania

91 5 07 089

Abstracting is permitted with the credit to the source. Libraries are permitted to photocopy beyond the limits of US copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 21 Congress Street, Salem, MA 01970. Instructors are permitted to photocopy isolated articles for commercial classroom use without fee. For other copying, reprint or republication permission, write to Director, Publishing Services, IEEE, 345 East 47th Street, New York, NY 10017. All rights reserved. Copyright © 1986 by the Institute of Electrical and Electronics Engineers, Inc.

FORTIETH ANNUAL FREQUENCY CONTROL SYMPOSIUM

Co-Sponsored by

**U.S. ARMY LABORATORY COMMAND
ELECTRONICS TECHNOLOGY AND DEVICES LABORATORY**

and

**THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.
ULTRASONICS, FERROELECTRICS AND FREQUENCY CONTROL SOCIETY**

SYMPOSIUM CHAIRMEN

General Chairman.....	Dr. John R. Vig, U.S. Army LABCOM
Technical Program Chairman.....	Dr. Leonard S. Cutler, Hewlett-Packard Company
Finance Chairman.....	Dr. Thomas E. Parker, Raytheon Research Div.
Publicity Chairman.....	Dr. Raymond L. Filler, U.S. Army LABCOM

TECHNICAL PROGRAM COMMITTEE

DAVID ALLAN National Bureau of Standards	JOHN KUSTERS Ball Corporation
DR. ARTHUR BALLATO U.S. Army LABCOM	THEODORE LUKASZEK U.S. Army LABCOM
MARTIN BLOCH Frequency Electronics, Inc.	DR. THRYGVE MEEKER Bell Laboratories
DR. JANET BROWN Fisher Controls International, Inc.	DR. THOMAS PARKER Raytheon Research Division
DR. LEONARD CUTLER Hewlett-Packard Company	VINCENT ROSATI U.S. Army LABCOM
DR. LAWRENCE DWORSKY Motorola, Inc.	LAUREN RUEGER Applied Physics Laboratory
DR. ERROL EERNISSE Quartex, Inc.	STANLEY SCHODOWSKI U.S. Army LABCOM
MARVIN FRERKING Rockwell International	DR. RICHARD SYDNOR Jet Propulsion Laboratory
DR. HELMUT HELLWIG Frequency and Time Systems, Inc.	DR. SAMUEL STEIN Ball Corporation
DR. WILLIAM HORTON Piezo Technology, Inc.	DR. JOHN VIG U.S. Army LABCOM
CHARLES JENSIK Piezo Crystal Company	DR. JOSEPH WHITE Naval Research Laboratory
DR. REYNOLD KAGIWADA TRW-ESD	DR. GERNOT WINKLER U.S. Naval Observatory
	DR. NICHOLAS YANNONI U.S. Air Force/RADC

TECHNICAL SESSION CHAIRMEN

PLENARY SESSION

Eduard Gerber, US Army LABCOM (ret.)

RESONATOR PROCESSING

John Kusters, Ball Corporation

FILTERS

Lawrence Dworsky, Motorola, Inc.

RESONATOR STUDIES I

Arthur Ballato, US Army LABCOM

ATOMIC FREQUENCY STANDARDS I

Robert Vessot, Smithsonian Astrophysical Observatory

RESONATOR STUDIES II

Errol EerNisse, Quartex, Inc.

ATOMIC FREQUENCY STANDARDS II

Richard Sydnor, Jet Propulsion Laboratory

FUNDAMENTAL PROPERTIES OF NATURAL & SYNTHETIC QUARTZ

Joel Martin, Oklahoma State University

OSCILLATORS I

Marvin Frerking, Rockwell International

FUNDAMENTAL PROPERTIES OF QUARTZ & OTHER CRYSTALS

John Vig, US Army LABCOM

OSCILLATORS II AND GPS NOISE

Albert Benjaminson, Systematics General Corporation

SAW DEVICES

Thomas Parker, Raytheon Company

FREQUENCY SYNTHESIZERS I

Victor Reinhardt, Hughes Space & Communications Center

SAW DEVICES AND OSCILLATORS

Janet Brown, Fisher Controls International

FREQUENCY SYNTHESIZERS II

Victor Reinhardt, Hughes Space & Communications Center

TRANSDUCERS AND SENSORS

Roger Ward, Quartex, Inc.

MICROWAVE AND MILLIMETER WAVE SOURCES

Rolf Weglein, Hughes Aircraft Company

SPECIFICATIONS & MEASUREMENTS

Robert Smyth, Piezo Technology

RESONATOR STUDIES III

Terry Flanagan, JAYCOR

TABLE OF CONTENTS

IN MEMORIAM

Jerrold R. Zacharias.....	1
Eduard A. Gerber.....	2
1986 AWARD WINNERS AND REMARKS.....	3
REMINISCENCES OF EARLY FREQUENCY CONTROL ACTIVITIES IN HONOR OF THE 40TH ANNIVERSARY OF THE FREQUENCY CONTROL SYMPOSIUM	
Plenary Session Remarks..... Eduard A. Gerber, U.S. Army Electronics Technology & Devices Laboratory (ret.)	8
Quartz Crystals Paved the Way..... Willie L. Doxey	9
Origin and Influence of the A F C S..... Virgil E. Bottom	15
Looking Back..... James C. King, Sandia National Laboratories	16
Early Frequency Control Reminiscences..... Marvin Bernstein	18
The Year When Quartz Ran Short..... Robert Adler, Zenith Electronics Corporation	20
The Frequency Control Symposium - Then and Now..... Warren Smith, AT&T Bell Laboratories	22
The Atomichron — 30 Years Ago..... Richard T. Daly, Quantronix Corporation	24
PROPERTIES OF QUARTZ AND BERLINITE	
Computer Modeling of Point Defects in Quartz..... T.M. Wilson, L.E. Halliburton, M.G. Jani, and J.J. Martin, Department of Physics, Oklahoma State University	26
A Radiation Growth Study of Acoustic Loss Related Defects in Alpha-Quartz..... J.J. Martin, H. B. Hwang, and T.M. Wilson, Department of Physics, Oklahoma State University	32
Characterization of Brazilian Lascas from Various Regions and Their Use for Synthetic Quartz Growth PART I. Lascas Study..... Hideo Iwasaki, Fumiko Iwasaki, Carlos K. Suzuki, Virginia A.R. Oliveira, Daniele C.A. Hummel, and Armando H. Shinohara, UNICAMP, Brasil	39

Characterization of Brazilian Lascas from Various Regions and Their Use for Synthetic Quartz Growth PART II. Properties Correlation.....	47
Carlos K. Suzuki, Armando H. Shinohara, and Virginia A.R. Oliveira, UNICAMP, Brazil Simão Takiya and Jozsef Kiss, ABC-Xtal Microelectrônica S.A., Brazil	
On the Disagreement in the Order of Magnitude of the Electroelastic Constants of Alpha-Quartz.....	54
Carl K. Hruska, Piezoelectricity Research Laboratory, York University, Ontario, Canada	
Infrared and Laser Spectroscopic Characterization of Aluminum Defects in Cultured Quartz.....	63
Herbert G. Lipson, Rome Air Development Center	
Developmental Results for the Production of High Quality Quartz.....	70
J.F. Balascio, Motorola, Inc. and A. F. Armington, Electromagnetic Materials Technology Branch, Solid State Sciences Division, Rome Air Development Center	
Experimental Study and Numerical Simulation of Quartz Crystal Etched Figures.....	76
C.R. Tellier, N. Vialle, and J.L. Vaterkowski, Laboratoire de Chronométrie, Electronique et Piézoélectricité, E.N.S.M.M., France	
Further Results on the Use of Surfactants in Chemically Polishing Quartz Crystals.....	86
Ronald J. Brandmayr and John R. Vig, U.S. Army Electronics Technology and Devices Laboratory (LABCOM)	
Quartz Analogues.....	91
John H. Sherman, Jr.	
Study of Irradiation Effects in Quartz Crystals Using Low-Temperature Dielectric Relaxation.....	96
S. Ling and A.S. Nowick, Henry Krumb School of Mines, Columbia University	
Berlinite: Characterization of Crystals with a Low Water Concentration and Design of Bulk Wave Resonators.....	101
J. Detaint, A. Zarka, B. Capelle, Y. Toudic, J. Schwartzel, E. Philippot, J.C. Jumas, A. Gouffon, J.C. Doukhan, France	
The Influence of Surface Finish and Metallization on Electrode Electromigration in Alpha-Quartz During Sweeping.....	115
John G. Gualtieri and Donald W. Eckart, U.S. Army Electronics Technology and Devices Laboratory (LABCOM)	
Impurities Migration Study in Quartz Crystal Resonators by Using Electroelastic Effect.....	121
R. Brendel, Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S., France and J.J. Gagnepain, J.P. Aubry, Compagnie d'Electricité et Piézoélectricité, France	
Evaluation of Mechanisms for Low-Dose Frequency Shifts in Crystal Oscillators.....	127
Terry M. Flanagan and Roland E. Leadon, JAYCOR and Darwin L. Shannon, TRW	
Low and Medium Dose Radiation Sensitivity of Quartz Crystal Resonators with Different AL-Impurity Content.....	134
J.J. Suter and R.H. Maurer, The Johns Hopkins University, Applied Physics Laboratory	
RESONATOR STUDIES	
ψ Angle in a Triply Rotated Cut, Determination and Control.....	140
Henri Méricoux and Jean-Francois Darces, Laboratoire de Cristallographie et Synthèses Minérales Faculté des Sciences et des Techniques, Université de Franche-Comté, France	

Dynamic Permittivities and Resistivities of the Equivalent Network Representing Plate Resonators.....	145
Arthur Ballato, Elizabeth Hatch, Muhammad Mizan, and Theodore J. Lukaszek, U.S. Army Electronics Technology & Devices Laboratory (LABCOM)	
Initial Stress Field and Resonance Frequencies of Incremental Vibrations in Crystal Resonators by Finite Element Method.....	152
P.C.Y. Lee and M.S.H. Tang, Department of Civil Engineering, Princeton University	
A Transmission Line Matrix Model for AT Quartz Thickness Shear Devices.....	161
Lawrence N. Dworsky, Motorola, Inc.	
Nonlinear Electroelastic Equations of Wave Propagation and Vibrations in Quartz Bars.....	168
M. Cengiz Dökmeci, Faculty of Aeronautics and Astronautics, Istanbul Technical University, Istanbul	
Three-Dimensional Finite Element Solution of the Lagrangean Equations for the Frequency-Temperature Behavior of Y-Cut and NT-Cut Bars.....	179
Y.K. Yong, Department of Civil/Environmental Engineering, Rutgers University	
Suppression of Anharmonic Spurious Modes by Modified Electrode Design Using Charge Cancellation.....	187
Hitoshi Sekimoto and Toshiyuki Ihara, Faculty of Technology, Tokyo Metropolitan University, Tokyo, Japan, Hozumi Nakata and Masaaki Miura, Nihon Dempa Kogyo Co., Ltd. Tokyo, Japan	
Variational Analysis of GT Cut Quartz Crystal Resonators with the Supporting Portions at the Ends.....	193
Hirofumi Kawashima, Seiko Electronic Components, Japan	
An Overtone-Mode Assigned At-Cut Crystal Resonator.....	201
K. Hirama, T. Shoji, and Y. Tanaka, Crystal Division, Toyo Communication Equipment Co., Ltd. Japan	
Overtone Response of Composite Bulk Acoustic Resonators.....	206
J. Rosenbaum and H.L. Salvo, Jr., Westinghouse Defense & Electronics Center, and S.V. Krishnaswamy, Westinghouse R&D Center	
TRANSDUCERS AND SENSORS	
A Quartz Fluid Density Sensor Pressure Transducer.....	211
Roger W. Ward and Errol P. EerNisse, Quartztronic, Inc.	
A Resonator Temperature Transducer with No Activity Dips.....	216
Errol P. EerNisse and Robert B. Wiggins, Quartex, Inc.	
Temperature Sensor Using Quartz Tuning Fork Resonator.....	224
Toshitsugu Ueda, Fusao Kohsaka, Toshio Iino, Daisuke Yamazaki, Yokogawa Hokushin Electric Corporation, Japan	
Double-Ended Tuning Fork Quartz Accelerometer.....	230
William J. Kass and Gary S. Snow, Sandia National Laboratories	
An Economical Touch Panel Using Saw Absorption.....	237
Robert Adler and Peter J. Desmares, Zenith Electronics Corporation	
SAW DEVICES; FILTERS	
Random Walk Frequency Fluctuations In Saw Oscillators.....	241
T.E. Parker, Raytheon Research Division	

Real Time, Interactive Saw Filter Computer Aided Design and Analysis Implementation.....	252
S.M. Richie, C.D. Bishop, and D.C. Malocha, University of Central Florida, Department of Electrical Engineering and Communication Sciences	
On a Continuous Representation of the Mode Shape in Acoustic Surface Wave Resonators.....	257
H.F. Tiersten and D.V. Shick, Department of Mechanical Engineering, Aeronautical Engineering & Mechanics, Rensselaer Polytechnic Institute	
An Analysis of the Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Supported Along the Edges.....	262
D.V. Shick and H.F. Tiersten, Department of Mechanical Engineering, Aeronautical Engineering & Mechanics, Rensselaer Polytechnic Institute	
Low Loss, Highly Stable Saw Devices on Quartz.....	269
T.N. Oliver, D.E. Bower, and J. Dowsett, STC Components Limited, Quartz Crystal Unit, United Kingdom	
Wideband Timing Tank Filters for Digital Transmission Systems.....	275
S. Yamamoto, T. Gounji, and J. Shimizu, Fujitsu Limited, Japan	
Acoustic Charge Transport Principles and Performance.....	285
M.J. Hoskins, M.J. Brophy, J.M. Dallesasse, M.J. Miller, J.W. Peterson, Electronic Decisions Incorporated	
Fast Channelizer — A New Frequency Sorting Technique.....	292
J.H. Elliott, R.B. Stokes, and K.H. Yen, TRW Electronic Systems Group	

SPECIFICATIONS & MEASUREMENTS

Group Delay Measurements — A Sensitive Method For Detecting Spurious Crystal Resonances.....	295
Ferdinand K. Euler, Rome Air Development Center, Solid State Sciences Division	
RF Spectrum of the Oscillator Signal Under Non-Stationary Phase Instabilities.....	300
Berni Joss, Laurent-Guy Bernier, and Fred Gardiol, Laboratoire d'Electromagnétisme et d'Acoustique Ecole Polytechnique Fédérale, Switzerland	
Aging Measurements on Quartz Crystals in the Batch Mode.....	306
E. Hafner and H.W. Jackson, XOTEX Corporation	
The Precise Determination by an Automatic System on the Resonance Frequencies of the Quartz Crystal Resonator.....	313
Yoshimasa Oomura and Yasuaki Watanabe, Tokyo Metropolitan University, Faculty of Technology, Japan	
Report on the Workshop on Traceability of Quartz Measurements to U.S. Standards.....	323
John A. Kusters, Ball Corporation, Efratom Division	

OSCILLATORS

A Frequency Adjustable Ultra-Compact, High-Performance Quartz Crystal Oscillator and Its Simple Temperature Compensation Method.....	325
O. Ochiai, F. Tamura, and Y. Mashimo, Seiko Electronic Components, Ltd., Japan	
Low Noise Crystal Oscillators Using 50-OHM, Modular Amplifier Sustaining Stages.....	329
M.M. Driscoll, Westinghouse Defense and Electronics Center	
Minimum Sideband Noise in Oscillators.....	336
J.K.A. Everard, King's College, London	

Digital Temperature Compensation of Crystal Oscillators Using Temperature Switches.....	340
Zeljko Aleksic and Dragan Vasiljevic, Faculty of Electrical Engineering, University of Belgrade, Yugoslavia and Aleksandra Pavasovic, Institute Mihailo Pupin, Yugoslavia	

A Crystal Oscillator with Bidirectional Frequency Control and Feedback ALC.....	344
A. Benjaminson, Systematics General Corporation	

Integrated Oven Controlled Quartz Crystal Oscillator.....	350
G. Marianneau, D. Hauden, and J.J. Gagnepain, Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S. associé à l'Université de Franche-Comté, France	

FREQUENCY SYNTHESIZERS

A Short Survey of Frequency Synthesizer Techniques.....	355
V. Reinhardt, K. Gould, K. McNab, and M. Bustamante, Hughes Aircraft Company, Space and Communications Group	

Army Frequency Agile Synthesizer Program.....	366
A. Bramble and J. Kesperis, U.S. Army Electronics Technology and Devices Laboratory (LABCOM)	

SOS Frequency Synthesizer Development.....	370
D.P. O'Rourke, RCA Communication and Information Systems Division	

Performance Analysis of the Numerically Controlled Oscillator.....	373
Elizabeth C. Kisenwether and William C. Troxell, HRB Singer	

MICROWAVE AND MILLIMETER WAVE SOURCES

The Coherence of a Radar Master Oscillator.....	379
R.D. Weglein, Hughes Aircraft Company	

The GAAS FET Oscillator — Its Signal and Noise Performance.....	385
Robert A. Pucel, Raytheon Co.	

Magnetically Tunable High Overtone Microwave Resonators.....	392
J.D. Adam and B.R. McAvoy, Westinghouse R&D Center and H.L. Salvo, Jr., Westinghouse Defense & Electric Center	

TIME COORDINATION AND DISTRIBUTION

Using Multiple Reference Stations to Separate the Variances of Noise Components in the Global Positioning System.	394
M.A. Weiss and D.W. Allan, National Bureau of Standards	

The U.S. Naval Observatory (USNO) PTTI Data Service.....	405
Gernot M.R. Winkler, U.S. Naval Observatory	

ATOMIC FREQUENCY STANDARDS

Active H-Masers for V.L.B.I. Applications.....	410
R. Barillet, P. Petit, J. Viennet, and C. Audoin, Laboratoire de l'Horloge Atomique, Equipe de Recherche du CNRS, associée à l'Université Paris-Sud, France	

A Hydrogen Maser at Temperatures Below 1K.....	413
Robert F.C. Vessot and Edward M. Mattison, Harvard-Smithsonian Center for Astrophysics, and Ronald L. Walsworth, Jr., Isaac F. Silvera, H.P. Godfried and C.C. Agosta, Lyman Laboratory of Physics, Harvard University	

A Microwave Pumped Cryogenic Hydrogen Maser.....	419
W.N. Hardy, M.D. Hürlimann, R.W. Cline, and A.J. Berlinsky, Department of Physics, University of British Columbia, Canada	
Single-State Selection System for Hydrogen Masers.....	422
Edward M. Mattison, Robert F.C. Vessot, Smithsonian Astrophysical Observatory and Shen Wei, Smithsonian Astrophysical Observatory and Shaanxi Observatory, Academia Sinica, People's Republic of China	
Progress Toward an Optically Pumped Cesium Beam Frequency Standard.....	428
R.E. Drullinger, Jon Shirley, D.J. Glaze, and L.W. Hollberg, National Bureau of Standards and A. DeMarchi, Istituto Elettrotecnico Nazionale "Galileo Ferraris" - Torino, Italy	
Effect of the Atom Transit Time on the Frequency Stability of Cesium Beam Frequency Standards.....	432
Claude Audoin and Vincent Candelier, Laboratoire de l'Horloge Atomique, France, and Jacques Vanier, Division de Physique, Conseil National des Recherches, Ontario, Canada	
A Novel Cavity Design for Minimization of Distributed Phase Shift in Atomic Beam Frequency Standards.....	441
A. DeMarchi, Istituto Elettrotecnico Nazionale "Galileo Ferraris" - Torino, Italy	
A Comparison of Various Alkali Gas Cell Atomic Frequency Standards.....	447
J.C. Camparo and R.P. Frueholz, Chemistry and Physics Laboratory, The Aerospace Corporation	
A Rubidium-Crystal Oscillator (RbXO).....	452
W.J. Riley and J.R. Vaccaro, EG&G Frequency Products	
A Disciplined Rubidium Oscillator.....	465
A. MacIntyre and S.R. Stein, Ball Corporation, Efratom Division	
Subminiature Rubidium Oscillator Model FRS.....	470
Werner Weidemann, Ball Corporation, Efratom Division	
AUTHOR INDEX.....	474
SPECIFICATIONS AND STANDARDS GERMANE TO FREQUENCY CONTROL.....	475
PROCEEDINGS AVAILABILITY INFORMATION.....	477

IN MEMORIAM

Jerrold R. Zacharias

On July 16, 1986 Jerrold R. Zacharias died unexpectedly at his home in Belmont, Massachusetts in his 81st year. Dr. Zacharias' career spanned several areas of physics during his extraordinarily productive life. In the context of time and frequency control his work on the radio frequency spectra of atoms and his superb competence at experimental work led to the successful development of the first atomic clocks and their subsequent manufacture. During the era beginning in the mid 1940's, prior to the space-age, he foresaw the possibility of testing Einstein's theory of relativity with atomic clocks and began working on clock concepts that directly led to the present cesium devices. He was truly the godfather of the atomic clock.

Dr. Zacharias was born in Jacksonville, Florida in 1905. He entered Columbia University at the age of 17 and he received the AB in 1926, the M.A. in 1927, and his doctorate in 1932, all from Columbia. Subsequent to his degree, his work with Prof. I.I. Rabi at Columbia led to many very important publications on the fundamental properties of atomic nuclei and, at Rabi's suggestion, led to the eventual development of the first practical atomic clock.

In 1940 he went to the Massachusetts Institute of Technology to head the Radiation Laboratory's Division on radar transmitter components. He continued his wartime career by going to Los Alamos to serve as director of the engineering division of the Manhattan Project. He returned to M.I.T. in 1946 to head the Nuclear Science Laboratory and continue his pre-war investigation on the properties of atomic nuclei using molecular beams techniques, work that led to the first cesium beam atomic clock.

In the late 1940's Dr. Zacharias was called on to head or participate in a number of studies important to the national defense, including Project Lexington (1948 nuclear powered flight), Project Lamplight (1954, continental defense), and Project Lincoln, from which the Distant Early Warning (DEW) line to guard our northern frontier was conceived. He was associate director of this project in 1954, which grew to become M.I.T.'s Lincoln Laboratory.

Dr. Zacharias was a member of the President's Science Advisory Committee for nine years between 1952 and 1964. For his services to national defense he was awarded the President's Certificate of Merit in 1948 and the Department of Defense Certificate of Appreciation in 1955.

Dr. Zacharias, or "Zach" as he was known to students and staff, was an extraordinarily gifted teacher who could describe complicated and arcane aspects of quantum mechanics in well understood and accurate analogies. He was, first of all, an experimental physicist who could conceive and design experimental apparatus with a clear understanding of the engineering involved and contribute directly to projects so that they could be built efficiently and work effectively.

The backwardness of physics teaching in secondary schools and the lack of laboratory experience in the education of high school students prompted Dr. Zacharias in 1956, the year before Sputnik, to form the Physical Sciences Study Committee. This committee revised the program of physics teaching and included many sophisticated experiments that could be done with very easily obtained hardware. By 1957, the program involved eight schools, and the following year 360



schools. By 1979, more than 200,000 students in 5,000 schools were involved. The program was carried on in 1958, and later, by Educational Services, Inc. (ESI) with Dr. Zacharias as director for academic affairs, and grew to include other aspects of teaching at all grade levels. In 1967, ESI merged with the Institute for Educational Innovation to become the Educational Development Center, Inc. with Dr. Zacharias as its vice president and founding trustee. In 1984 he was honored by the International Commission for Physics Education for "long and distinguished service to physics education" and for being "a teacher of teachers."

In particular, we remember him at the 1986 presentation of the I.I. Rabi award "for technical excellence and outstanding contributions in the fields relating to atomic and molecular frequency standards." Other awards include the Oersted Medal of the American Association of Physics Teachers (1961) and the National Science Teachers Association Citation for Distinguished Service to Science Education (1969). He received honorary degrees from Tufts University, Oklahoma City University, St. Lawrence University, Lincoln University, and Brandeis University.

In 1966 M.I.T. named him Institute Professor, a faculty title reserved for colleagues of special merit and distinction. He retired from the faculty in 1970, continuing as Director of M.I.T.'s Educational Research until 1972.

He was a fellow of the American Association of the Advancement of Science, and the Institute of Electrical and Electronic Engineers, a member of the National Academy of Sciences, the American Academy of Arts and Sciences, the American Physical Society and the American Association of Physics Teachers.

Those who were fortunate enough to work with him will remember Zach's extraordinary gifts of leadership, the joyous way in which he immersed himself in his work, and his genius and intuition at seeing the core of situations. His high spirited and energetic way of life was contagious and continues to influence large numbers of his students and associates.

Dr. Zacharias is survived by his wife, Leona (Hurwitz), two daughters, Susan and Johanna, and three granddaughters.

A service of remembrance of Dr. J.R. Zacharias will be held at the Massachusetts Institute of Technology, Cambridge, Mass., in Auditorium 10-250 at 4:00 P.M. on Friday, October 24, 1986.

IN MEMORIAM

EDUARD A. GERBER

Dr. Eduard A. Gerber died peacefully in Bridgton, Maine on August 8, 1986 at age 79. He was born in Fuerth, Bavaria, Germany, on April 3, 1907, and received the M.S. and Ph.D. degrees in physics from the Institute of Technology in Munich, Germany, in 1930 and 1934, respectively.

In 1935, he joined the scientific staff of the Carl Zeiss Works, Jena, Germany, and was in charge of research and development in piezoelectric crystals. From the time of his arrival in the United States, in 1947, until 1954, he was crystal research consultant to the Signal Corps Engineering Laboratories at Fort Monmouth, NJ. From 1954 to 1961, he served as Director of the Frequency Control Division, U.S. Army Signal Research and Development Laboratory; from 1961 to 1963, he was Director of the Solid State and Frequency Control Division, U.S. Army Electronics Laboratories; and from March 1963, to 1970, he was Director of the Electronic Components Laboratory, U.S. Army Electronics Command, all at Fort Monmouth. Since his retirement from the federal civil service in 1970, he had been a consultant to the U.S. Army Electronics Command (now U.S. Army LABCOM).

Dr. Gerber was a Life-Fellow of the IEEE, cited for his contributions to piezoelectricity and frequency control. He was also a Fellow of the American Association for the Advancement of Science, a member of Commission I, U.S. National Committee of the International Scientific Radio Union, the American Physical Society, and the New York Academy of Sciences. He holds eight U.S. patents and three German patents and has published 38 professional papers and contributions to books. He was also coeditor of the two-volume monograph Precision Frequency Control.

Dr. Gerber received the Department of the Army Decoration for Meritorious Civilian Service in December 1965 and July 1970 and the C. B. Sawyer Memorial Award in 1981.

The foregoing is a purely one-dimensional scientific sketch of a warm and multi-dimensional human being who will be acutely missed by those fortunate enough to have known and worked with him. His was an enviable combination of optimistic faith, scientific curiosity, philosophic and literary depth, engaging demeanor, and evident good will.

As noted, Ed Gerber came to the U.S. in 1947, the year of the first Frequency Control Symposium. Since then he continually played a key, but largely unheralded, role in its nurturing and growth. He had a distinct preference for scientific and engineering work, but because of his management skills and ability to deal amiably and equitably with people, he rose through the administrative ranks. This diminished his scientific output but not the locus or intensity of his interests, particularly his greatly loved field of frequency control. And from his administrative position he



was able to encourage the development of the AFCS, and ancillary functions; e.g., he had been a member of the C. B. Sawyer Memorial Award committee since its inception in 1966.

On this Fortieth Anniversary year of the AFCS, he organized and chaired the Plenary Session devoted to "Reminiscences of Early Frequency Control Activities"-many of these activities he himself had helped to mold. And whether it was in the description of his own work, for example, showing the influence of crystal plate parallelism on mode spectrum purity (1943-44), or the recounting of stories of Sommerfeld's lectures, or of the Bechmann/Telefunken - Straubel/Zeiss AT-BT cut rivalry of 1933-34, Ed was filled with scientific enthusiasm, an enthusiasm and zest he never lost; but now we have lost him.

Dr. Gerber's last technical paper will appear in the October 1986 IEEE MTT Transactions, and is entitled "Advances in Microwave Acoustic Frequency Sources." This gentle man will not be forgotten; Frequency Control Symposium attendees have only to look about them, and in the Proceedings of each year.

1986 AWARD WINNERS

THE CADY AWARD

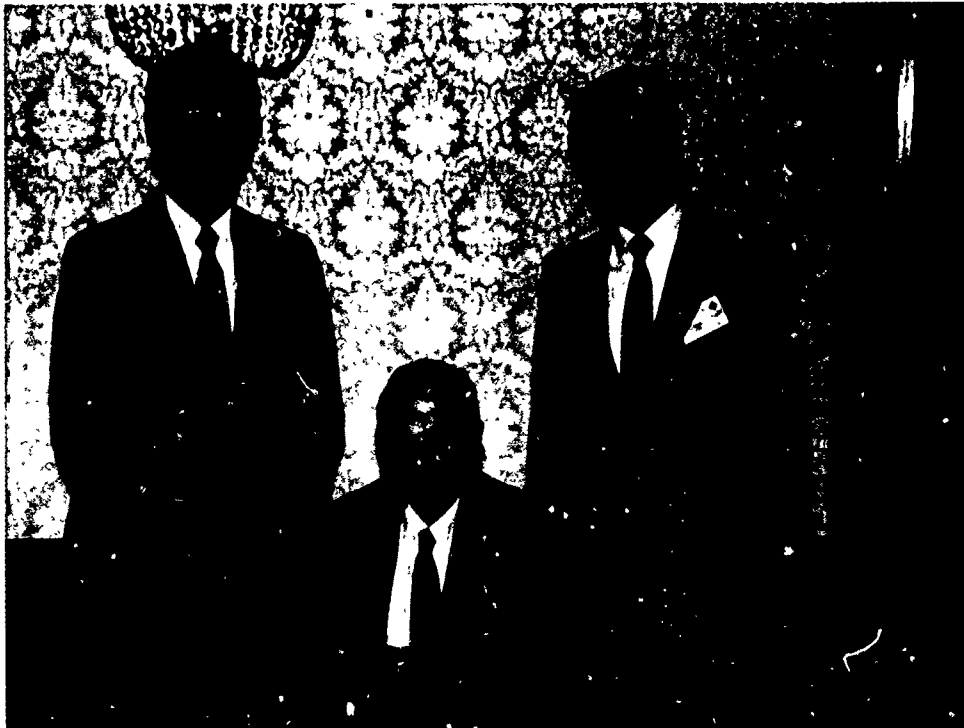
The Cady Award was presented to Juergen H. Staudte for his pioneering contributions to the photolithographic processing of quartz devices, especially the development and commercialization of quartz tuning forks for timekeeping. The award was presented by Roger Ward, Quartztronics, Inc.

THE RABI AWARD

The Rabi Award was presented to Dr. Jerrold R. Zacharias for his contributions to the development of atomic frequency standards, especially his scientific leadership, pioneering demonstration of the technology, and entrepreneurial initiative which led to the commercialization of atomic standards. The award was presented by Prof. Norman Ramsey, Harvard University.

THE SAWYER AWARD

The Sawyer Award was presented to Dr. Larry E. Halliburton for his contributions toward the characterization of cultured quartz using infrared absorption, electron spin resonance, acoustic loss, and thermoluminescence measurements. The award was presented by Dr. Eduard A. Gerber, US Army Electronics Technology and Devices Laboratory (retired).



Larry E. Halliburton, Sawyer Award winner; Jerrold R. Zacharias, Rabi Award winner; and Juergen H. Staudte, Cady Award winner, at the award dinner.

FORTIETH ANNUAL FREQUENCY CONTROL SYMPOSIUM

Award Banquet Remarks

Arthur Ballato

US Army Electronics Technology & Devices Laboratory
Fort Monmouth, New Jersey 07703-5000

Anniversaries are opportunities to look back and reminisce; this can be a salutary thing every once in a while, even for the practitioners of our modern science.

I would like to share with you a very condensed version of some less-discussed aspects of the growth and development of the AFCS, and we begin with what I shall call "The Internationalization of the Annual Frequency Control Symposia," shown in Fig. 1. I've plotted in the upper portion the percentage of papers from other than North America versus year. The first gathering occurred in a conference room at Fort Monmouth in 1947, and consisted of Signal Corps personnel and WW II contractors. The next three meetings were held at Gibbs Hall, the FM Officers' Club. In 1951, the 5th AFCS was moved to the Berkeley-Carteret Hotel in Asbury Park, where it stayed through the 13th. The first off-shore papers were given by members of the British Post Office. By 1956, when the first written proceedings were issued, off-shore participation was below 10%, but even then had begun the upward cosmopolitan swing that has continued to this day. The long-term trend line has slope of 80% per century, as you can see from the figure!

With the 14th, the symposium moved to Atlantic City, where it stayed for twenty years: first at the Shelburne, until the Silver Anniversary in 1971, and then at Howard Johnson's, save for the 27th which detoured to Cherry Hill due to fire. During the twenty years at AC, perhaps due to the salubrious salt air, there developed what I call the "Capistrano Effect." Many an attendee arrived right on schedule, despite the yearly change of date, and despite missing the announcement; they just homed in from all over. More about this in a bit. In 1966 the C.B. Sawyer Memorial Award was established.

During the AC years the symposium continued to mature; printing of the proceedings shifted from the FM print plant, with distribution gratis, to the EIA in 1969. In 1967-68 our industry was undergoing a period of introspection and self-evaluation. A biennial symposium was suggested. A survey of the industry was published in 1968. The annual symposium survived.

The move to the Philadelphia Marriott in 1980 was mainly occasioned by the advent of certain stochastic processes that were set in motion in AC at that time, with the realization by our hosts that frequency control engineers and scientists knew too

much about statistics and random variables. The 36th AFCS saw the introduction of the first registration fee, the transition to a conference management firm, and a concerted effort to consolidate the exhibit area. The next year brought co-sponsorship of the symposium between the US Army and the IEEE Sonics and Ultrasonics Group. Co-sponsorship has led to a further internationalization of the symposium, and to the transformation of IEEE G-SU into the IEEE Ultrasonics, Ferroelectrics, & Frequency Control Society.

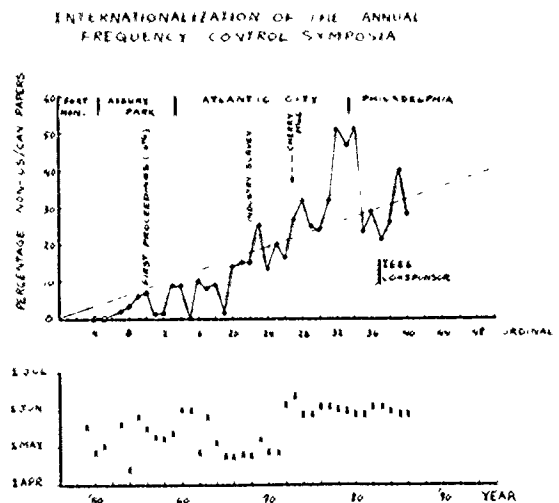


Figure 1. Internationalization of the Annual Frequency Control Symposia.

The lower portion of Fig. 1 graphs the temporal variations in the holding of our spring-time event. This plot shows good long-term stability, but rather poor Allan variance for sampling times of a year or two. By the mid-70s the AFCS became phase-locked with Memorial Day, and this continued through the change of venue to Philadelphia and up to the present time. For some years after the last move the AC Capistrano Effect continued to operate, and each year a few visitors would show up late in Philly with sand in their shoes and a question mark on their faces. Our attendees are nothing if not loyal!

It is interesting to consider some of the more

remote roots of the symposium as well as aspects of its enduring continuity. Prof. Cady, one of the giants of our field, and the man for whom one of our awards is named, was a student of Max Planck, and received his Ph.D. in 1900 in Berlin. He attended the 4th AFCS (26-27 April 1950) at age 76, and gave a paper entitled "Piezoelectricity as a Branch of Thermodynamics." When he attended the 18th AFCS in 1964 at age 90, he gave a short but prescient talk in which he observed the trend in certain branches of physics for the number of authors to approach the number of words in an abstract. The costs of crystal resonator experiments, by comparison, remain modest!

At the start, the AFCS was completely given over to what has been called "classical" frequency control: quartz resonators, the search for natural quartz substitutes, and oscillator measurements. "Atomic & molecular," or "quantum mechanical" papers were not long in forthcoming, however, and that brings me to my next topic. I recently found in my files an abstract from the 9th AFCS in 1955, the year before written proceedings were issued. This physics abstract has but one author, and it is none other than Prof. Zacharias of MIT, whom we honor tonight with the Rabi Award.* In all probability it hasn't been seen since 1955, so let us read what it says (Fig. 2). The frequency is only 230 Hz from the present definition, and its author is certainly prescient in respect to stability and lifetime.

SYMPOSIUM PROGRAM
NINTH ANNUAL REVIEW OF TECHNICAL PROGRESS
SIGNAL CORPS ENGINEERING LABORATORIES
FREQUENCY CONTROL BRANCH
FORT MONMOUTH, NEW JERSEY
25, 26, 27 May 1955
Hotel Berkeley-Carteret, Asbury Park, N. J.

AN ATOMIC FREQUENCY STANDARD
JERROLD R. ZACHARIAS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

The hyperfine structure resonance line of atomic Cesium at approximately 9192.632 mc/sec has been used to stabilize the frequency of a microwave signal generator. This resonance frequency, describable as the Larmor precession frequency of the valence electron of atomic Cesium in the magnetic field of the atomic nucleus, is observed by the atomic beam magnetic resonance method as a resonance curve with a half-width of approximately 200 cycles/sec. The techniques necessary to make it into a practical device have become available only in the last few years in connection with researches on rare isotopes of the alkali atoms. These techniques will be described.

Preliminary results show that the short-time stability thus obtained (for times less than one second) is better than 1 part in 10^9 and that the average stability for long times will be considerably better than this. The beam tube has run satisfactorily for periods as long as 50 hours indicating that a sealed-off tube with a life of several years should be quite possible.

NOTES

Figure 2. The first AFCS cesium paper, 1955.

I came to work at the Signal Corps (as we were called then), in 1958. I was fresh from MIT, with a liking for circuit theory learned from Prof.

Guillemin, a student of Arnold Sommerfeld. When I got to Fort Monmouth, Dr. Gerber† put me to work under Rudolf Bechmann, another of Sommerfeld's boys. He set me straight about the beauties of crystal physics -- inner space --, lessons I'll never forget, despite my occasional equivalent circuits.

Cady wrote me a letter in 1972 in which he said, "I hope you will continue your interest in the field of piezoelectricity," and apropos of that I have here (Fig. 3) my anniversary gift to the Annual Frequency Control Symposia.

ANNUAL FREQUENCY CONTROL SYMPOSIA				
YEAR	ANNIVERSARY	FORMULA	CRYSTAL CLASS	PIEZO (?)
1951	5 - WOOD	(CELLULOSE/LIGNIN)	FIBROUS CRYSTALLINE POLYMER	YES
1956	10 - TIN/ALUMINUM	Sn/Al	n/n m/m n/m	NO, NO
1961	15 - "CRYSTAL"	SiO ₂	ISOTROPIC	NO
1966	20 - CHINA	Al ₂ O ₃ ·2SiO ₂ ·2H ₂ O	POLYCRYSTALS (KAO-LIN)	NO
1971	25 - SILVER	Ag	n/m	NO
1976	30 - PEARL	CaCO ₃ (CALCITE/ARAGONITE)	POLYCRYSTALS (3m/3mm)	NO, NO
1981	35 - CORAL	CaCO ₃ (CALCITE)	POLYCRYSTALS (3m)	NO
1986	40 - RUBY	Al ₂ O ₃	3m (CORUNDUM)	NO
1991	45 - SAPPHIRE	Al ₂ O ₃	3m	NO
1996	50 - GOLD	Au	n/m	NO
2001	55 - EMERALD	Al ₂ (BeSi ₂ O ₆) ₃	6/m mm (BERYL)	NO
2006	60 - DIAMOND	C	n/m	NO

Figure 3. Piezoelectric AFCS anniversaries?

When I got to the 40th without finding one example of a single-crystal piezoelectric, what was there to do but --- extrapolate! My search, alas, was not rewarded with success; all of the crystals belong to the holohedral classes. Perhaps the system of anniversaries can be augmented with the addition of more interesting materials such as quartz (class 32), lithium tantalate (class 3m), and gallium arsenide (class 43m). In fairness, we should also include substances representative of the other branch of our symposium, such as cesium; I think that would be hyperfine! Thank you.

* We note with sorrow the passing of Prof. Jerrold Zacharias on July 16th, 1986, at age 81.

† We note with grief the passing of Dr. Eduard A. Gerber on August, 8th, 1986, at age 79.

THE SUCCESSIVE OSCILLATORY FIELD METHOD AND THE HYDROGEN MASER

Norman F. Ramsey

Lyman Physics Laboratory
Harvard University
Cambridge, MA 02138

INTRODUCTION

Since I have always been interested in the origins of new ideas in physics, I have decided to devote my time to a description of the origins of my two principal contributions to precise measurements of time and frequency -- the method of successive oscillatory fields and the hydrogen maser.

New ideas and inventions originate in a wide variety of ways but often they originate from the interaction between an incentive and a quite unrelated stimulus. This was true of both ideas that I shall discuss this morning.

THE METHOD OF SUCCESSIVE OSCILLATORY FIELDS

I obtained my Ph.D. from Columbia in 1940 on the basis of my work with I. I. Rabi on the then-new molecular beam magnetic resonance method. After several years of interruption by World War II and after brief periods teaching at Columbia and heading the Brookhaven Physics Department, I came to Harvard University in 1947 and obtained some ONR support for constructing a molecular beam magnetic resonance apparatus to be 10 times more accurate than any previous such apparatus. In principle I could do this by lengthening the radio-frequency transition region from the previous 15 cm length to 150 cm which would thereby increase the transition time by a factor of ten and, in accordance with the Heisenberg Uncertainty Principle, narrow the observed resonance by the same factor. But I knew that in some past efforts to narrow the resonance in this fashion, the width of the resonance had increased rather than diminished. The reason was that it was difficult to keep such a long magnetic field region at a uniform value and when the field varied the resonance was at different nearby frequencies and thereby appeared to be broadened rather than narrowed.

I had hoped to overcome this problem with a magnet design which should have provided a more uniform magnetic field, but I was having difficulty in achieving the uniform field. My worry that I might not meet my objective provided the

incentive for wanting to have a new idea.

The stimulus suggesting the new idea came from a quite different direction -- from a lecture I was then giving on the Michelson stellar interferometer in my course on physical optics. When I was a student in Cambridge, P. I. Dee told me a rather dramatic way in which to remember the Michelson stellar interferometer. If one were on the verge of being able to tell if a star were double or single, and if there were sufficient light intensity, he could double the resolution of the telescope and thereby resolve the question merely by taking a can of black paint and painting over all of the telescope lens except for two narrow slits at opposite ends of a diameter. Not only would this double the resolution, but clearly the image would be unaffected by the quality of the glass hidden behind the paint. I immediately thought that it would be wonderful if I could do something analogous to make the molecular beam magnetic resonance pattern independent of the quality of the magnetic field. However, the analogue is not a perfect one and it took me some hours after class to realize that, if I used two separate but coherently driven oscillatory fields so that the molecules went through one first and then the other at the end the magnetic resonance region, the resonance would be narrowed and would not be broadened by field irregularities since the resonance frequency would depend only on the average magnetic field between the two resonance coils. Even after recognizing the possibility, I only slowly came to appreciate the many other advantages the method had, including the possibility for overcoming first order Doppler shifts, of working at wave lengths comparable to or smaller than the separation of the two coils, and of selecting only slow or long-lived atoms which would therefore give resonances narrower than those expected from a simplistic use of the Heisenberg uncertainty principle.

THE ATOMIC HYDROGEN MASER

The incentive for the hydrogen maser invention was my disappointment and even jealousy in having missed what seemed to be a beautiful idea for achieving much higher precision in atomic frequency measurements. This was Jerrold Zacharias'

invention of a fountain experiment, in which a slow beam of molecules would be directed upward and allowed to fall back down under gravity. Since approximately a second would be required for such a reversal of direction under gravity, any resonance observed during that time should have the unprecedentedly narrow width of 1 Hz. Although the experiment never did succeed because of the unexpected absence of the required ultra-slow atoms from the atomic source, the proposal was of great value since it stimulated the invention of several other successful experiments, including the hydrogen maser and our bottled neutron experiments.

The stimulus for the specific idea of the hydrogen maser came from my quite unrelated work in high energy particle physics. I was at that time chairman of the joint Harvard-MIT committee responsible for the construction and operation of the 6 GeV Cambridge Electron Accelerator which required a large underground circular ring tunnel several hundred feet in diameter. There are many worries involved in the responsibility for such a project including the possibility of the accelerator not working. What could one do to extract some good physics from such a tunnel if the accelerator failed? One possibility was to make a long circular molecular beam all the way around the tunnel. However, if inhomogeneous magnetic fields were used to constrain the molecules in the necessary circular orbit, the resonance frequencies would be greatly shifted. On further thought it became clear it would be better to confine the atoms to a hollow circular tube with a suitable internal wall coating. However, in such a tube the frequency of collisions with the sidewall was primarily dependent on the smallest width of the tube and was approximately independent of its length so that storage in a closed box would be just as good and just as feasible as storage in a long circular ring. My graduate student, Dan Kleppner and I then developed what we called the "broken atomic beam" experiment for cesium using such a coated storage bottle. Although the experiment was successful, only a few hundred wall collisions occurred before the atoms were lost. By then we realized that the light atomic hydrogen should stick on the wall of the containing vessel for shorter times and should thereby have a smaller wall shift of its frequency. But hydrogen was hard to detect with the means then available so we calculated the possibility of detecting the electromagnetic radiation. When this calculation proved to be favorable, we realized that we could observe spontaneous maser oscillators by continuously feeding in state-selected atomic hydrogen with only the high hyperfine state being selected and the atomic hydrogen maser was born.

40th Annual Frequency Control Symposium - 1986

REMINISCENCES OF EARLY FREQUENCY CONTROL ACTIVITIES IN HONOR
OF THE 40th ANNIVERSARY OF THE FREQUENCY CONTROL SYMPOSIUM
= = PLENARY SESSION REMARKS = =

Eduard A. Gerber

U.S. Army Electronics Technology & Devices Laboratory (ret.)
USALABCOM, Fort Monmouth, NJ 07703-5302

It seems to be very appropriate to look back once in a while into the past to ponder what had happened and what it can teach us for the future. This is valid for an individual life as well as for organizations, such as our frequency control community. Obviously, any anniversary offers a special opportunity to look back and reminisce. Thus, the Program Committee has scheduled a session for this purpose and invited outstanding men of slightly advanced middle age to speak shortly of the past and, thus, help us to celebrate the 40th Anniversary of our Symposium.

However, a feeling of sadness is intermixed with our elation. A man who certainly would be with us today is no more among the living. Roger Sykes passed away on February 16, 1986, one day short of his 79th birthday. The frequency control community lost one of its most prominent members and I personally a dear friend. I would like to ask you to stand up for a moment to honor his memory. Thank you.

Our symposium started in 1947 in a conference room of the Squier Signal Laboratory in Fort Monmouth. It was attended only by personnel from the three services, contractors working on frequency control problems and members of a subpanel of the old Research and Development Board.

The purpose of this first Symposium was to review progress on the various contracts and assist the military in future program planning. I will never forget this interesting, but very modest beginning. During the next 3 meetings the Symposium was expanded to include others and went to Gibbs Hall, also located in the Ft. Monmouth area. Due to increasing attendance it moved to Asbury Park in 1951 and stayed there until the first meeting in Atlantic City in 1960, the 14th Symposium. Finally, the Symposium was held for the first time in Philadelphia in 1980 and has stayed there henceforth. Another milestone in the history of our conference is the fact that IEEE has been cosponsoring the Symposium since 1983 which certainly adds to its importance and prestige.

At the celebration of the 25th Symposium in 1971, Roger Sykes and I had the gratification of presenting an overview of achievements in the crystal field based on earlier Symposium papers, and of listening to Prof. Ramsey who spoke on the history of atomic and molecular frequency control. Today, we have again the pleasure of having Prof. Ramsey with us. We will enjoy listening to him and other distinguished old-timers who will speak about their past experiences and recollections. The sequence for our speakers will follow as much as possible the history of the Symposium and the development of our technology.

QUARTZ CRYSTALS PAVED THE WAY

By Willie L. Doxey

May 1986

Early Military Communications

Means of communication available to support military operations, prior to the invention of electrical means -- by Samuel F. B. Morse -- were dependent upon personal messenger service. The communication had to be written, spoken or read; and it had to be delivered and received by the proper person without interference from outsiders. It had to be received over great distances and precisely timed for proper arrival.

Further, if verbal or written, it was essential that contents be precisely spelled out or, if previously coordinated with the receiver, be so carefully "garbled" that only the person receiving would thoroughly understand it.

A number of other means of communicating effectively made use of sound and light. Some of these systems included torches, pillars of fire by night, clouds by day, mirrors flashing in sunlight, pistols and other pyrotechnics; the semaphore, which Napoleon developed into a system of 1200 stations between Paris and Moscow; and of course, sirens, gongs, whistles, and voice. Pigeons were a primary means of delivering messages, that were used extensively during World War I and on special assignments during World War II.

Technological Advances in Military Communications

Advances in technology, resulting in signal transmission electrically over wire and cable, revolutionized communications making available to the military reliable, instant, secure communications for tactical operations by use of telephone, telegraph, and teletype. Further technological advancements added electrical signaling to that of wire and cable devices, to include radio, radio-telephone, and teletype along with radar, radiosondes for weather reporting, radio control of airways, and numerous other applications.

These significantly enhanced the capabilities of the military forces -- tactically, logistically, and administratively. These advancements in technology rapidly led into the development of tactical equipment in the fields of communication, radar, countermeasures, meteorological equipment, all-weather surveillance devices (among others) which enhanced the Army's capabilities in all areas of warfare. As part of the Army Service Forces, the Army Signal Corps took its place as one of three major groupings of men and materiel -- the other two being the Army Ground Forces and the Army Air Corps.

The Signal Corps Mission

The Signal Corps described its mission as "not only (the provision of) communications facilities for the War Department in the zone of Interior and all overseas theaters, but (also) the design, procurement, construction, installation, and major maintenance for radio, radar, telephone and telegraph communications for operations of the U.S. Army throughout the world".

Early Technology Advancements

Along with technology advancements in electrical/electronic devices and equipment, parallel developments were proving successful in heavier-than-air vehicles. The Signal Corps immediately sensed a need for air-to-ground communication and took action to establish requirements for necessary equipment and facilities.

Signal Corps Specification No. 486 dated 23 December 1907 was signed by James Allen, Brigadier General, Chief Signal Officer of the Army. A contract placed with the Wright Brothers, 10 February 1908, was signed by Charles S. Wallace, Captain Signal Corps, US Army, and Orville Wright (trading as Wright Brothers), 1127 West Third Street, Dayton, Ohio. The contract (total cost of \$25,000) was contained on one legal size page and the specification on one-and-one-half regular size pages for a heavier-than-air vehicle to be delivered on or before 28 August 1908 (6 months and 18 days).

The flying speed of the machine was required to be at least 40 miles per hour in still air, with a fuel load for 125 miles and two persons (approximately 350 lbs.). A most interesting and unique feature of this contract was the "positive and negative incentive clause" as follows:

40 mph - 100% of cost; 39 mph - 90%;
38 mph - 80%; 37 mph - 70%; 36 mph - 60%.
Below 36 mph - rejected.
41 mph - 110% increased in increments of
10% to 140% at 44 mph.

It is also interesting to note that sixty years later a similar "positive and negative incentive clause" found its way into a major contract - and may still be used in certain contracts.

The Signal Corps responsibilities for airplanes were short-lived -- with the establishing of the Army Air Corps in early 1930. However, the development of communications and other airborne electronics remained with the Signal Corps for several years.

Some Mobilization Activities Prior to US Entry into WW II

Mobilization planning and military build-up in the 1940-1941 era were receiving very high priorities. Military aggression by the Rome-Berlin Axis, with well-trained troops equipped with modern weapons, showed complete success with little or no significant resistance.

The overrun of Ethiopia by Mussolini and Hitler's invasion of the Rhineland (without any resistance) caused grave concern among our Allies, and in fact all members of the League of Nations. The annexation of Tokyo to the Rome-Berlin Axis removed all doubts in the minds of nations nearest the danger -- such as England, France, Canada, and the United States -- that "Military Build-Up" is the top priority item of the country.

One example of a Congressional response to the critical needs was to increase the 1939 authorization of two million dollars for "Signal Corps Educational

orders to procure critical components and communications equipment" to 35 million dollars by the end of the fiscal year. For example, the 1939 mobilization plan for dry batteries (an essential and critical component) was 1.5 million which was revised to 5 million for fiscal year of 1939. It turned out that this expanded estimate fell short of actual requirements since production of dry batteries exceeded 21 million annually during the 1940-41 years. Signal Corps total obligations went from 6 million dollars in 1939 to 3 billion dollars in 1943. (These estimates included some requirements for our allies, namely Canadians and British).

Technology Growth and Impact on Communications

During the two years (1939-1941) prior to "Pearl Harbor", significant technological advancements were made in electronics and communications which made possible a series of new types of radio sets that more completely satisfied the needs expressed by the Armored Force. These radios were multichannel receiver/transmitter sets that would permit instant "pushbutton" tuning for selection of pre-set frequency channels. They made use of the piezo-electric quartz crystal, the most stable method for controlling radio-frequency (RF) oscillators known at that time -- and still without peer for tactical military radio equipment.

The Research Laboratories at Fort Monmouth, New Jersey, (Coles Signal Laboratory, known originally as Eatontown Signal Laboratory) sponsored the development of prototype models for field testing, type-classification, and standardization.

The original models of these FM (frequency modulation), short range radio sets were built by Fred Link in late 1939. Models of these equipments were similar to the commercial radio sets furnished to local police departments by "Link Radio Co". The Signal Corps provided the Armored Force with several models of these FM radios for field testing.

Armored Force communication personnel were impressed with the performance of the FM radios. Further demonstrations to commanders of the Armored Force resulted in command decision to switch from AM (Amplitude Modulation) to FM (Frequency Modulation) for all short-range armored tanks. This decision led to substantial orders for the FM radios.

Paralleling the efforts in frequency modulation, engineers in the Coles Signal Laboratory had been working with Western Electric, among other commercial contractors (along with in-house efforts), to develop a new radio set to replace the tuned R-F circuits with quartz crystals. This turned out to be the initial step to introduce quartz-crystal control of radio frequencies in combat radio sets.

Design of the First Armored Force FM Radio Set

Paralleling these efforts, in late 1939 and '40, the Bell Laboratories and Western Electric Co. had the gigantic task of redesigning the AM prototypes of Armored Force Type III transmitter/receiver sets into FM "pushbutton" "crystal-controlled" sets to be identified as Signal Corps Radio SCR-500 series and used for mobile army communications.

This project was among the 10 highest priority projects in the Signal Corps at this time. It had the support of the Armored Force which gave top priority for communication equipment to support ongoing maneuvers in the United States. These actions constituted irreversible commitment to the role of the quartz crystal for frequency control in combat military communication and electronic equipment.

The prime concerns of those opposed to the conversion were (a) delay in delivery of communication equipment to the Armored Divisions presently on maneuvers and (b) the imminent danger of becoming involved in major conflict without communications. The position of officers and engineers supporting conversion were essentially twofold: (a) production was currently underway for Link FM radios to equip the First and Second Armored Divisions with communications and (b) Western Electric Co. was scheduled to begin production in late December 1941 on the new 80 channel (with 10 preset channels on any of 80 channels) FM, pushbutton, crystal-controlled radio sets known as Signal Corps Radio SCR-500 series.

Perhaps the most convincing support was the statement made by Colonel James D. O'Connell (who subsequently became Chief Signal Officer) when he said repeatedly, "Yes, we can give you radios; but with quartz crystals -- we can give you communications". The Chief Signal Officer felt some relief with ongoing production of the Link FM radios for delivery of 2,500 sets to meet the requirements of the First and Second Armored Divisions which were already engaged or scheduled for maneuvers within the United States; and by the fact that Western Electric would be shipping FM sets in December.

The Field Artillery, having observed the performance of the Western Electric design of the SCR-500 series FM radios, placed a priority order on the Signal Corps to redesign the SCR-500 series radios to operate in the frequency band assigned to the Field Artillery. This redesign extended the frequency band to include 120 channels (in lieu of 80) and resulted in the new Signal Corps Radio SCR-600 series which also included crystal-control, pushbutton tuning with 12 preset channels on any selected frequency.

The Field Artillery gave top priority for delivery of these FM radio sets to be delivered in late 1941 and for Laboratory tests to be completed in February 1942. Plans included delivery of 600 sets to field units in 1942.

FM Radios for the Infantry

Commitments by the Armored Force and Field Artillery to FM crystal-controlled radios, forced action to include the Infantry in the family of FM radios. Several types of experimental radio sets were designed and constructed for use by the combat soldier; but for reasons of size, weight, and power requirements were considered unsuitable.

Infantry requirements for a man-pack set to operate satisfactorily over a 7 to 9 mile range could not be developed under a twenty-five pound restriction. After reducing the requirement to a range of two miles and increasing the weight to 35 pounds, two companies attacked the program using standard AM circuits; but they continued to have difficulties meeting user requirements.

These Infantry requirements were made known to Paul Galvin and Dr. Dan Noble of Motorola who immediately began design work on the smallest possible radio set using Dr. Armstrong's early development of FM technology and the state-of-the-art in miniature vacuum tubes. The efforts of Motorola and Philco fulfilled the Army's dream for a radio "transceiver" that a soldier could carry in one hand and operate while walking.

The radio weighed five pounds complete with battery. It was single frequency, crystal-controlled, and provided communication over a two-mile range. The radio took the nomenclature of Radio Set SCR-536 but soon became known service-wide as the "handle-talkie".

Production began in late 1942 while the Infantry requirements for these radio sets ballooned into the millions. This added new and critical requirements for quartz crystals, on an already overloaded crystal manufacturing industry, which had not yet gotten "off the ground". This additional crystal requirement had further profound impact on the critical supply of raw quartz available for production of the added millions of quartz crystals.

Airborne VHF Radios

Another radio set, the airborne VHF command set, Radio Set SCR-522, was developed by the Airborne Radio Laboratory (ARL) at Wright Field, Dayton, Ohio, which was initially under the direction of the Signal Corps Labs. The airborne VHF radio was also crystal-controlled, with pushbutton tuning of four preset channels.

The SCR-522 used the basic design of the British VHF radio being produced for the British in the United States. It had preset crystal-controlled channels, and operated in the frequency band of 100 to 156 megacycles.

Design and production of test models of this new airborne command radio set were accomplished by the Bendix Radio Corporation. The service and acceptance tests were performed under direction of Aircraft Radio Laboratory at Wright Field, Dayton, Ohio.

Requirements for Quartz Crystals Continued to "Snowball".

The decision of the War Department to provide the Army with new communication equipment in astronomical quantities was indeed a bold step. New production sources were required for ruggedized components for all of the FM and VHF equipment. Among these components, without doubt, the "QUARTZ CRYSTAL" required the most immediate and concentrated attention. The four (4) Signal Corps radio types listed here were among the chief crystal "gobblers":

SCR-500 series - Armored Force - 10 preset channels - 80 channels - 2 crystals for each channel (1 receiver and 1 transmitter)

SCR-600 series - Field Artillery - 12 preset channels - 120 channels - 2 crystals for each channel (1 receiver and 1 transmitter)

SCR-522 - Airborne Command Set VHF - 100-156 megacycles - 20 kc channel spacing - preset channels

SCR 536 - Infantry - 30-40 megacycle band

There were many other Army radio sets which used

quartz crystals (some in relatively large numbers) as well as Navy, some commercial requirements, as well as for airborne communication and test equipment.

The number of quartz crystals required for the initial production planned for the Signal Corps radios listed above simply "dwarfed" the total production capacity of the crystal industry. (This total industrial capacity, as informally estimated by a few old-timers, was at most a few tens of thousands of units per year). The projection of full mobilization requirements for crystal production, drew sharp focus on the critical and urgent problems of expanding the crystal industry. The crystal requirements for mobilization escalated from tens of thousand per year to tens of million per year.

The Critical Supply of Raw Quartz

Signal Corps commitments to crystal control of tactical radio communication equipments produced an immediate surge in the increase of contractual orders for quartz crystal units. These increased contractual requirements resulted in a critical and in-depth analysis of (a) the available raw quartz suitable for producing crystal units, and (b) the amount of quartz crystals required for total mobilization. Quartz crystals were added immediately to the list of strategic materials along with tungsten, tantalum, steatite, and a few others. The analysis further resulted in (a) priority funding for additional stockpiling of quartz, and (b) educational orders for improving production facilities for increased efficiency for orientation, sawing, lapping, and finishing of the crystal blanks.

Signal Corps planners estimated the need for procuring 300,000 pounds of quartz crystals for the fiscal year of 1942. However, no one held hopes for getting delivery for several reasons:

- 1) The United States had authorized (and funded) the stockpiling of 106 thousand pounds of quartz in 1939 and less than half had been delivered.
- 2) All radio-grade quartz that the United States had access to came from Brazil. Efforts by the US Geological Survey and the Signal Corps to locate domestic sources were complete failures.
- 3) Japan had taken the greatest part of Brazilian exports, while the British were next highest. In 1941, Japan increased their imports from Brazil by a factor of four.

Through Metal Reserves Corporation, the Reconstruction Finance Corporation (RFC) was charged with building up the United States' supply of quartz both for stockpiling and for production. This expediting of quartz import resulted in increases in the quantity of quartz shipped but a drop-off in the quality of material.

Quality of Quartz Required for Military Radios

Quartz that did not have at least one natural face could not be used by the industry since X-ray diffraction equipment had not yet been developed. Also, material should be free (or relatively so) from optical and/or electrical twinning in order to be used for frequency control plates.

To improve the quality of quartz imported to the United States, the Signal Corps trained six

technicians and provided inspection facilities for shipment to Brazil to inspect all of the quartz prior to shipment to the United States. This inspection resulted in immediate improvement in the quality of raw quartz for crystal manufacturers and for the government stockpile.

Several actions were taken to increase and/or stretch the supply of raw quartz which continued to become increasingly critical, due primarily to the increased requirements for the production of quartz crystals. These steps were to:

- 1) Instruct manufacturers to use smaller size pieces of raw quartz. Due to production efficiency, in prior years, many manufacturers discarded crystals weighing less than 30-50 grams. (One manufacturer recovered several hundred pounds of crystal which he had used to pave a portion of his driveway!)

- 2) Accelerate the programs with General Electric and North American Philips to develop X-ray machines for accurate orientation of both faced and unfaced quartz. (This added a substantial quantity of previously unusable quartz to the production lines -- as well as improved production techniques.)

- 3) Increase production of quartz from the mines in Brazil.

The raw quartz problem was among the top ten items on the Signal Corps Priority One list. The following actions were taken:

- 1) In an effort to alleviate the immediate shortage ("and get the Signal Corps off our backs"), Transportation authorized the "hold" of one ship to be filled with several thousand pounds of quartz and started towards New York. A U-Boat sank the ship about one day after it left the harbor!

- 2) That devastating blow to our quartz supply generated a second action which also found its way to "Davy Jones' locker". This second action was to assemble and ship mechanized mining equipment (such as bulldozers, tractors, and diggers) to Brazil to increase production from some of the more productive mines. But the U-Boats got this ship, also. Prior to purchase and shipment of replacement equipment, an analysis of mine locations, lack of roads, along with logistic support and operation of the heavy equipment, resulted in cancellation of the replacement order.

- 3) Another effort to increase mine production was to offer "incentives". Accordingly, the price per kilo was increased (substantially) nearly double. However, the life style and habits of the Brazilian miners (Garimpos) were not fully understood by the United States' negotiators. This approach 'or a brief period, actually resulted in a net loss. The miners seemed to enjoy the life style in the village more than that of the mines; and their habits were to collect the money from their mining and "live-it-up" in the village until their money was exhausted. Only then did they find it necessary to return to the quartz mines.

The supply of quartz was so critical at one time in 1942 that commercial DC-3 aircraft were employed to fly some quartz into the United States, at a reported cost of two dollars per pound. At this time (1942), the amount of quartz in the stockpile was unofficially reported as ten thousand pounds (although historical reports indicated that 50 thousand pounds should have been delivered). Only one request for release of quartz from the national stockpile was

received by the Quartz Crystal Section, OCSigO. This was for 200 pounds of "large" optical-grade quartz to fabricate special lenses for a classified research program. This urgent request was received by messenger on Easter Sunday, 1942, and after numerous telephone calls, release was obtained by a signature from an office in the White House Annex. Further withdrawals from stockpile were discouraged.

Redesign to Reduce Size of the Quartz Crystal

The one single problem which had to be solved, (and promptly) before production on the new radio equipment could begin, was redesign of the quartz oscillator plate to reduce the size. The quartz plate as originally designed and fabricated for frequency control was one inch square. The amount of raw quartz of "radio grade" (that is, free of optical or electrical twinning, bubbles, or mineral deposits) in all of Brazil could not possibly meet the millions of crystal units required for military radio equipment using a one-square-inch quartz plate.

Engineers from the Signal Corps Laboratory and Bell Telephone Laboratory, along with engineers from equipment manufacturers, redesigned the crystal to be the frequency control element and not a power-generating element. The vast majority of applications in military equipment made use of quartz plates about three tenths of an inch square.

This reduction in size required an extra stage of amplification in the circuitry of the radio sets. These changes had the immediate effect of increasing the supply of usable raw quartz by several hundred percent. This reduction in size of the quartz element actually made possible the success of the Signal Corps program for the best combat radio equipment for all military services.

Signal Corps Leadership in Mobilization

The Signal Corps commitment to design, produce, and provide the Armored Force, Field Artillery, Infantry, and Air Corps with this new combat radio equipment -- using the new technology of FM, VHF, and quartz-crystal tuning in the face of near certain involvement in a world-wide conflict -- was heavily laced with "impossible" problems. There were expressions of "doubt", "fear of failure", "impossible task". Some of these were loud and from high levels. However, they were short-lived.

The strong leadership of Signal Corps officers, namely Brigadier General Roger B. Colton and Colonel James D. O'Connell, and their aggressive actions in mobilizing the giants in the electronics industry, as well as the middle-sized and small (one-horse) companies, began to show positive changes. The spirit of competition was (almost overnight) turned into genuine cooperative attitudes of open-house sharing of ideas, experiences, and trade secrets.

A key move of the Chief Signal Officer was to reassign General Colton to a key position in the Office of the Chief Signal Officer (OCSigO) with responsibility and authority to do "that which was necessary" to meet the Signal Corps commitments to provide the Army with combat communications equipment superior to that of the Axis, and in the shortest period of time possible. Among the early actions taken by General Colton was to reassign Colonel James D. O'Connell to the General Development Division of OCSigO and, in addition to his other duties, to make him responsible for OCSigO staffing of the total crystal program. In February-March of 1942, Colonel O'Connell established the Crystal Coordinating Section

in General Development Division in Building Temp "A" located on Buzzard's Point adjoining the Army War College in Washington, DC.

This writer was included in the original group of eleven civilians and two officers. By June of 1942, the section had increased to more than 40 people which included physicists, engineers, mineralogists, and geologists as well as support personnel.

This section provided coordination and staffing of any or all matters pertaining to expansion and operations of the quartz crystal industry to meet the needs of the War Department. This included the Navy, and later the US Air Force and our allied forces as required. It included responsibility to assure adequate supply of raw quartz and priority allocation as and when required.

Development of Crystal Saving Circuitry

The Signal Corps could not ignore the possibility that its only source of raw, radio-grade quartz (from Brazil) could be interrupted. This potential tragedy prompted the Signal Corps to initiate a back-up development program for experimental radio sets that provided frequency selection using only one crystal unit -- "Crystal Saving Circuitry".

A priority contract was placed with Zenith Radio Corporation (which confirmed discussions between Signal Corps Laboratory engineers and Zenith personnel), for the design and production of three multichannel FM transmitters (including six receivers) which used a single quartz crystal to stabilize frequencies of four master oscillators. Target date for delivery of the transmitters and receivers was three months. Research on master oscillators using a single crystal was being done at the Signal Corps Laboratory along with certain commercial manufacturers. Results of this research were made available to Zenith.

The transmitters and receivers were delivered on schedule. Testing and evaluation could not be completed until late 1942. At this date, multiple contracts had been placed with several prime contractors, each of whom had placed subcontracts with dozens of vendors for components, materials, and supplies and deliveries were supporting production lines already shipping radio sets to Armored Force troops on maneuvers and in training. (Sets were being shipped without a full complement of crystals.) However, the minimum number of crystals for 12 channels were adequate until the crystal manufacturers could be expanded.

The War Department priority for communication equipment in support of mobilization efforts would not tolerate any (repeat any) delays for changes that would stop or even momentarily interrupt production of communication equipment.

The Signal Corps guidance from the CSO (Chief Signal Officer) level was simple and without ceremony: Revive the crystal saving program for the next generation of radio sets after we have fulfilled our commitments to the Armed Forces by delivering the best radio sets that could be produced by any country -- (and complete with a full complement of crystals).

Field Crystal Grinding Teams

Domestic maneuvers and war games simulating combat conditions revealed a vital need for security in radio communication. This required the capability of over-night changing to entirely new frequencies in

order to successfully carry out certain mission assignments. It required changing the frequency of a number of crystal units, in a matter of hours or overnight, depending upon the magnitude of the operation or assignment.

The changes required special equipment and highly-trained, skilled personnel to provide such a delicate and vital service. The Signal Corps Laboratories provided the required service -- a crystal grinding team, along with equipment and supplies, to be assigned as an integral, mobile part of a division or company headquarters unit under control of the commander's chief of staff.

A mobile, crystal-grinding facility consisted essentially of the following: One each two-and-a-half-ton truck (enclosed) complete with self-contained power generator, special electronic test and measuring equipment; abrasive-type lapping, finishing equipment, and receiving necessary supplies; crystal blanks finished to 10 kc spacing over the frequency band of the equipment; and all materials and supplies to sustain operations for a 6 to 9 month period.

Many assignments could be met by changing frequencies of existing crystals to slightly higher frequency. The European theater (including the United Kingdom) was provided with six complete mobile crystal grinding units, each consisting of one Officer, three enlisted technicians, and each with a full complement of equipment and supplies.

The two major crystal depots, Philadelphia and Sacramento, in addition to storage and distribution of crystals, were equipped with a crystal fabricating facility where large orders for special crystal units on discrete frequencies were processed for special tactical missions. These special crystal fabricating facilities were also used for training of crystal technicians for the teams as well as specialized crystal technicians for crystal manufacturers.

The US Air Force established Crystal Bank organizations with headquarters in Wright-Patterson Air Force Base, Dayton, Ohio. Two regional crystal banks, one at Newark and the other at San Francisco, each had a "specialized fabricating" facility for finishing and testing special crystals for classified tactical operations. These two crystal grinding facilities services each of the 15 Air Forces except the one in the China-Burma-India theater, which was serviced by a special team established in New Delhi. Each of the 15 Air Forces had its own crystal bank and was serviced by either the Newark or the San Francisco regional bank.

These Army and Air Force specialized fabricating facilities, along with a specialized quick-reaction procurement activity with continuing blanket procurement orders from several crystal manufacturers did indeed fulfill the numerous requirements necessary for successfully accomplishing vital tactical missions during the entire war.

Aftermath of the Big War

I vividly remember these famous words of President Roosevelt in an address to the American people shortly after the attack on Pearl Harbor: "This is the war to end all wars". No one could ever have predicted that these United States would be involved in what turned out to be another major military conflict, in less than a decade -- the "Korean War".

We, as a nation, had made outstanding progress in the conversion of our economy and life style from a

total support of military operations to that of civilian needs and desires. We had effectively disassembled the most powerful total war machine ever built (moth-balling as much as appeared practical) while disposing of the remainder in the most effective and practical way possible.

Manufacturing facilities producing war materials were closed immediately, contracts canceled and either dismantled or converted to the production of civilian products. Production of quartz crystals for military use fell to near zero. Actions were taken to stockpile quartz blanks, some crystals, and of course, all raw quartz was stockpiled or sold to a few remaining crystal manufacturers.

Mobilization efforts for the Korean conflict were not comparable to the problems which confronted us in 1941-44. However, we found scores of the same problems, with the same equipment, and facing high priorities, allocation of materials, recalling some of the previous manufacturers into production and expanding existing production facilities.

Most of the communication equipment was the same as used in World War II. Our problems were primarily those of expanding existing facilities and arranging priorities and procurement orders.

The nucleus of scientific and engineering personnel were still available; and expanding technical work forces was manageable. However, there was much-much turbulence in expanding the industrial base for crystal production to meet the requirements for the first 18 months of conflict.

IN SUMMARY

Expansion of the crystal manufacturing industry began early in 1942. Those few manufacturers in existence were supported with educational funds and were referred to manufacturers of radio equipment requiring crystals for further support. Manufacturers of civilian commodities were rapidly converted to war material through an established manufacturer or referred to the Signal Corps Laboratories for orientation.

Perhaps, the most effective efforts were those of prime contractors of the radio equipment, such as, Western Electric who built and/or enlarged its present facility to meet their own crystal requirements; and Motorola/Galvin who established and supported a family of producers who met their production needs.

Beginning in 1942, the Signal Corps sponsored production forums which met periodically (monthly at first) in conveniently located cities for discussion of technical problems associated with production of crystals. These meetings were held throughout the war years.

After 1945, the Signal Corps R&D Laboratory along with the Air Force, Navy, commercial laboratories, National Bureau of Standards, and universities established an annual technical meeting. Called the "Frequency Control Symposium", this symposium still provides a platform for world-wide exposure of accomplishments and problems in the field of frequency control and related technologies.

Research and Development has contributed to the complete solution of numerous problems, military and civilian, which have faced us over the decades. One of the more critical ones mentioned here has been completely eliminated - that of Piezo electric

crystals for Frequency Control. With slightly more than one million dollars of Army R & D Funds, industry has progressed to the point where the US is completely free from natural quartz for the production of Frequency Control Devices. More than 90 percent of crystals produced now makes use of synthetic or "cultured" quartz crystals.

ORIGIN AND INFLUENCE OF THE A F C S

Virgil E. Bottom

During World War II many scientists and engineers left their positions in universities and industry to contribute their skills and knowledge to the immediate task of bringing the war to a successful conclusion. This was true in the field of applied piezoelectricity and technology. As the war came to an end late in the year 1945, most of us made plans to return as soon as possible to the positions we had left three or four years earlier.

It was then that a decision was made by the military authorities which proved to have far-reaching consequences. That decision was to keep the scientists and engineers actively working in the fields in which they had gained experience in the military laboratories. The mechanism chosen for implementing this decision was to award Research and Development Contracts to the institutions from which they had come.

During the years of the war research and development work on quartz crystal units was done at the Signal Corps Laboratories at Fort Monmouth, at the Naval Research Laboratories in Washington, and at the Aircraft Radio Laboratories of Wright Field at Dayton, Ohio. After the war most of the work on quartz crystal units was assigned to the Signal Corps Laboratories including the supervision of the R & D contracts.

A considerable number of R & D contracts were made with various institutions which had contributed staff members to work on the development of quartz crystal units. It soon became apparent that some coordination was needed among the various groups working in the field. It was impractical for the small staff of the Quartz

Crystal Group at Fort Monmouth to visit all of the contractors so arrangements were made to have them come to the Laboratories once each year to report on their work. In this way, each R & D group could report, not only to the sponsoring agency, but to each of the other groups working in the field. These meetings ultimately grew into the Annual Frequency Control Symposium which continues today.

One of the objectives in setting up the program of R & D in quartz crystal technology was to keep a group of workers active and abreast of the field in case the need for their services should again develop. For forty years it has done this. But it has done much more. Our knowledge and understanding of the quartz crystal unit has been greatly extended through the efforts of those who were first introduced to the device through their work during the war. But, more importantly, a number of younger men and women were introduced to the subject of piezoelectricity and its applications through these R & D contracts. Many of them are leaders in the field today and their contributions far exceed those of us who were their teachers.

So, starting as a conference of R & D contractors, the Annual Frequency Control Symposium has grown to become the foremost gathering in the world for the dissemination of information in the field of theoretical and applied piezoelectricity. Today it attracts workers from many countries who meet annually to discuss their ideas and to exchange information in the field of frequency control. The progress made in the field in the past forty years is due, in no small part, to the Annual Frequency Control Symposium.

LOOKING BACK

James C. King
Sandia National Laboratories
Albuquerque, New Mexico 87185-5800

I am reminded of the man who said he didn't regret growing old -- it would be a pleasure to have time to reminisce about events of his youth. The trouble was when he finally reached that time of life, he couldn't remember a thing.

Fortunately, the organizers of this session caught a few of us with memory intact--or nearly so. Unfortunately, what I know of the early days of frequency control and these symposia may not go back far enough in the past to be of historic significance. I trust that other speakers in this session can recall the very beginnings of this event and thus satisfy the curiosity of the historians among us. I shall caution you though: my recollections are personal, and for this I ask your indulgence.

When I joined Roger Sykes' department at BTL 33 years ago, I felt like a Johnny-come-lately. At first I shared an office with Art Warner who gave me the impression that the technology of precision frequency control needed only a bit of tidying up here and there. The AT-cut 5-MHz 5th overtone resonator was threatening to replace Jack Griffin's 100-kHz GT-cut resonator as the Bureau of Standards' frequency standard. But Griffin's crystals were tough to beat and Jack put up a valiant last-ditch effort. I was privileged to witness the kind of performance one never forgets--when Jack Griffin adjusted the Q and the resonant frequency of the GT-cut crystal at the same time by polishing the edges of the plate. Jack was a Picasso with etchants, polishing compounds, and cleaning agents.

In retrospect, joining Sykes' group when I did was a lucky coming together of people, places and events. At that time Sykes' group had some of the top resonator and oscillator design engineers in the business--Irv Fair, Warren Smith, Art Warner, Jack Griffin, Larry Koener, and others. Also, A. C. Walker at BTL and Danforth Hale of Clevite Electronics had only recently succeeded in growing quartz crystals large enough to fabricate resonators. In the meantime, Warner had developed the planoconvex 5 MHz resonator for measuring the intrinsic Q of quartz.

I arrived on the scene at this point with a background in low temperature physics and physical acoustics. Within a year after joining BTL, I had the very good fortune to meet Dr. Ed Gerber of the Signal Corps Engineering Laboratory.

Ed was very kind and showed a real interest in my study of the conduction process in quartz. A few months after we met I became the principal investigator of a Signal Corps contract with Dr. Erich Hafner as the contract monitor. (This was Erich's first contract-monitoring job.) I was to study the fundamental properties of quartz. The contract imposed a discipline I was unused to--namely: "You must submit a quarterly report of accomplishments"--and an implicit requirement: Present a report of your most recent work at the Annual Frequency Control Symposium. (I got the definite impression that my presentation had better be good or no contract extension.) And so, the first year I participated--I think it was 10th Annual Symposium--the program committee in a display of uncommon tolerance allowed me time to discuss the initial results of my contract work. Ray Mindlin was session chairman. He put me at ease right away by saying in his introduction that he didn't understand my abstract, and what I was about to say was unlikely to change matters in any case.

But I pressed on and reported on some rather unremarkable measurements of the axial conductivity of quartz. The audience was neither large nor particularly responsive. Not long after this performance, Roger Sykes proposed that he would continue to study the macroscopic behavior of resonators while I would do well to devote myself to the microscopic properties of quartz. I wish he had used the term "atomistic" instead of "microscopic," but I suspect he had something else in mind. For the next 6 years I immersed myself in my "microscopic" studies, courtesy of the Signal Corps Engineering Laboratory.

During these 6 years the competition from within the Engineering Lab at Fort Monmouth became quite stiff. Erich Hafner was developing precision instrumentation to measure the electrical characteristics of high-frequency crystal resonators. Don Hammond designed and built a very sensitive double-crystal goniometer to measure the parameters of quartz lattices. Andy Chi was beginning his painstaking study of differences in the frequency-vs-temperature characteristic curves for quartz of different origins. And Art Bollato began his study of the force-frequency effect in circular plates.

My contract with the Signal Corps allowed me a remarkable degree of freedom, and this 6-year period proved to be the watershed of my career. Out of it came the quartz sweeping process, high-growth- rate hi-Q quartz, hi-Q resonators for operation above 400°C, lithium-doped hi-Q quartz, and rad-hard quartz. I'm sure that many of you here today can recount similar experiences through your association with the sponsors of this symposium.

The technology of frequency control continues to challenge the skill and ingenuity of our best engineers and scientists. And the demand for precise time measurement seems insatiable. Without question these symposia provide a unique forum for people active in the field to disclose the results of their work and to invite constructive criticism. We owe a great debt to the founders of the Annual Symposium on Frequency Control and to their successors. I am honored to join you in celebrating the 40th anniversary of this event.

EARLY FREQUENCY CONTROL REMINISCENCES

Marvin Bernstein, Retired
1137 Hope Road
Asbury Park, NJ 07712

This paper is a brief description of some developments in frequency control undertaken during the period of 1942-1946 during the Second World War while employed at the Signal Corps Laboratories at Fort Monmouth, NJ.

This paper will deal with my early experiences with projects at Fort Monmouth involving some aspects of frequency control. This work ranged from low frequency magnetostriction oscillators to quartz crystal units and test sets. For the most part, however, the assigned tasks concerned developments in the field of crystal testing.

In early 1942, the use of frequency modulated radios operating about 30 Mhz required very large numbers of CT face shear crystals but the technique for attaching wires to the surface of the quartz was not fully developed at this time. For this reason, a project was established to see if magnetostriction resonators could supplant the quartz units and by 1943, a practical resonator was developed. At this time, Western Electric was producing vast numbers of face shear crystals and the magnetostriction device was not used except as a calibrator for a long range radar equipment.

The small group working on the substitute for low frequency crystals was then transferred from the Radio Branch to the Crystal Branch of the Signal Corps. Since there were many problems in the testing of crystals, this Group began to work in this area. A Radio Set was in production that used a series resonant unit but there were no series resonant test sets. The first units were tested at parallel resonant in a typical oscillator circuit with an off-set frequency specified. This obviously did not work well and means for testing these new units had a high priority.

The first test set used a stable variable LC oscillator and the crystal was placed in a Pi network. Tuning of the oscillator frequency was varied until maximum transmission was observed on an output indicator. This frequency was not that of series resonance but represented the minimum impedance of the crystal. This test set was useful but not very practical for production use.

Capt. C. J. Miller, Branch Chief, in 1943 discussed the possibility of using a Colpitts type oscillator with the crystal unit inserted in series with the inductor of the LC circuit. This unit was assembled and tested using a calibration resistor mounted within an empty crystal holder to simulate the stray capacitance of a crystal. The results of these tests showed that this assembly was indeed accurate and was put into use in the production of CR-8/u crystals.

During one period of the War, crystals were tested using a replica of the using oscillator. In the late part of the War testing was done using one of two test sets. Widely used was the CES-1 test set

which consisted of a Pierce oscillator, with variable capacitors in both the grid and plate parts of the circuit. These capacitors were set to values that resulted in the crystal under test operating at the proper frequency and with a minimum value of rectified grid current that allowed the unit to operate in the using radio set. The other test set was the AN/TSM-1 which was a Miller type oscillator and operated in about the same fashion.

The greatest problem during this whole period was the difficulty in keeping these oscillators properly correlated since changes in tube characteristics greatly affected the units. Personnel of the Quality Assurance Section were on the constant move, carrying along a "STANDARD" set and trying to keep the set of the manufacturer adequately correlated. Late in the War, work began to develop a test set that did not require this periodic readjustment but would be capable of adequate accuracy without external calibration.

With the success achieved with the CR-8/u test set, work began on a new type of instrument patterned on the series resonant circuit. This set would be designed to operate over a wide frequency, load capacitance and power level range.

The CR-8/u crystal test set was successful in that the only calibration required was the use of a non-inductive resistor to set both the frequency of operation and the passing level of equivalent series resistance for the unit under test. The only reason that this circuit could not be used directly for the new set was the bad effect that this substitution resistor stray capacity had on the circuit.

It was known that the impedance at the center of the coil would be very low and thus the inductor of the oscillator circuit was split into two equal parts. A switch was provided at this point so that either the crystal or substitution resistors could be inserted into the inductor of the Colpitts oscillator. The circuit Q was reduced by means of resistors from each side of the switch to ground to avoid high frequency parasitic oscillation. This small change in the circuit resulted in the test set that would be called a C.I. Meter.

Two members of the Crystal Branch, in 1944, that had encouraged the development of the C.I. Meter were C. B. Davis and A. Pritchard. Mr. Davis was the Army member of the Standardization Group that established the series of crystal units in Mil-C-3098. A new test set was absolutely needed for this specification. Mr. Pritchard spent much time and effort both in suggesting ways to increase the accuracy of the set and to arrange for tests to be made at the Bureau of Standards to determine the resulting accuracy.

With further development of the basic circuit, test sets were manufactured that covered the

frequency range from 70 Khz to 140 Mhz. These series of instruments were used as the specified test set in Mil-C-3098 for many years.

THE YEAR WHEN QUARTZ RAN SHORT

Robert Adler, Fellow, IEEE

Zenith Electronics Corporation
Glenview, Illinois 60025

ABSTRACT: A 1 MHz magnetostrictive resonator for frequency control is described. It was developed at the start of World War II when quartz crystals were in short supply, but the shortage ended before it went into production.

* * * *

This is the 40th anniversary of the Frequency Control Symposium. It might be logical to assume that the history of frequency control began in 1947. What I would like to tell you about happened several years earlier; you might think of it as prehistoric events.

When this country entered World War II in December 1941, quartz crystals for frequency control purposes had been in use for about 20 years, but they were not a mass production item. Suddenly they were needed everywhere. The company I had just joined--Zenith--started building a Signal Corps receiver (SCR 808) which used a single 1 MHz crystal to provide a frequency reference ("whistle stop tuning") for about 100 VHF channels separated by 100 KHz. This receiver, of course, was only one of many that were suddenly being produced. Add to this some thousands of small transmitters being built all over the country, and you can imagine the result: a severe shortage of quartz crystals developed. Quartz came from South America; synthetic quartz was not yet available, and processing facilities were few.

Early in 1942, one of our engineering executives came to me. "We will need a lot of 1 MHz quartz crystals", he said, "and we can't get them. Is there any chance that you can come up with a substitute?" At that particular time, I was excluded from taking a direct part in the frantic efforts of the engineering department because, having arrived from Austria not too long before, I was classed as an enemy alien. In accordance with regulations, the company had tucked me and a few others away in a building remote from the main engineering department, in a couple of nice, quiet rooms we lovingly referred to as Siberia.

I went to the library and read up on magnetostrictive oscillators. People had built some pretty stable oscillators at frequencies of about 50 KHz, using nickel alloy rods about 5 cm long that vibrated in their lowest longitudinal mode. A frequency of 1 MHz would call for a rod 2.5 mm long, which seemed too short to be practical. In a quartz crystal, frequency is often determined by the thickness, i.e. the smallest dimension. With magnetostriction in metal this is not practical because of eddy currents. In a conventional magnetostrictive oscillator, frequency is determined by the length of the rod, i.e. its largest dimension. I wondered whether perhaps the width, i.e. the intermediate dimension of a thin, flat body would be a good choice.

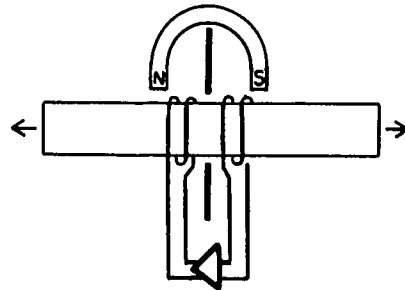


Fig. 1: Magnetostrictive rod oscillator (about 1930).

I therefore started experimenting with washers --flat rings of thin sheet metal with inside and outside diameters of about 6 and 11 millimeters

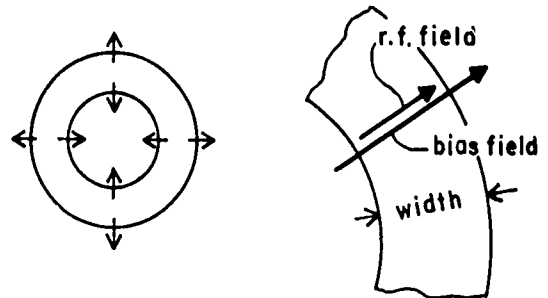


Fig. 2: Magnetostrictive washer vibrating in width extensional mode. Bias and r.f. field are both radial.

respectively. The frequency-determining dimension was the radial width: In the desired mode, the inside and outside contours would move in opposite directions. Two flat pancake coils, one on each side, were used to couple to the radial r.f. magnetic field produced by the vibrating washer. A structure resembling a small speaker magnet provided the radial bias field (Fig. 3 next page).

The washer was made of Elinvar, a Ni-Cr-Fe alloy specially prepared by Arnold Engineering in Marengo, Illinois. Elinvar is designed to have a Young's modulus which changes very little with temperature. A minor modification produced a low temperature coefficient of frequency. Note that Ni-Span-C, developed after the war, is very similar to Elinvar.

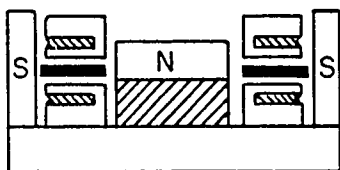


Fig. 3: Assembly of biasing magnet, washer (black) and pancake coils for width extensional mode.

We soon found that a particular shear mode was even more stable with temperature. Shear modes in

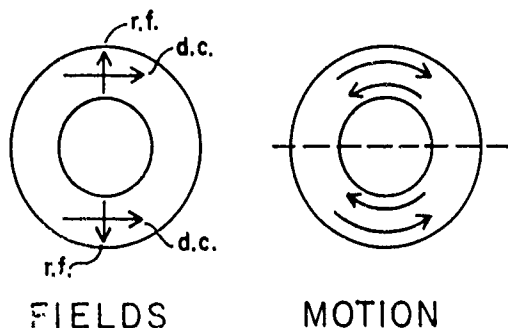


Fig. 4: Magnetostrictive washer vibrating in split shear mode (one nodal line). Bias field is unidirectional, r.f. field radial.

magnetostrictive materials are produced when bias field and r.f. field are crossed; the preferred shear mode included a nodal line across the center of the washer. Because of the lower shear velocity,

the washer had to be smaller than before: Its inside diameter was now 4 mm and its outside about 7 mm. To keep the nodal line parallel to the direction in which the metal had been rolled, two positioning notches were added to the center hole.

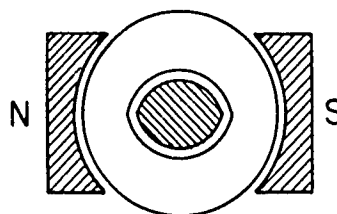


Fig. 5: Assembly of biasing magnet and notched washer (split shear mode). Coils are not shown.

We punched out quite a few of these washers, annealed them, -- they all came out within a few KHz--and adjusted them to frequency by hand, using sandpaper.

The quartz crystal for which the receiver had originally been designed fit into an octal tube socket and used only 3 of the 8 contacts. With the help of a few capacitors and resistors connected to the remaining contacts, we made the oscillator tube work interchangeably with the quartz crystal or with its magnetostrictive replacement--the "tin crystal", as it was nicknamed by Commander Mac Donald, founder and then still head of Zenith.

I would like to tell you now that the tin crystal was a roaring success, that millions were produced and that it helped win the war. But that is not what happened: Just when it was ready, the floodgates opened, quartz crystals started arriving by the thousands and the little washers were swept away. And all this happened five years before the first Frequency Control Symposium.

The Frequency Control Symposium - Then and Now

Warren Smith
AT&T Bell Laboratories
Allentown, PA

INTRODUCTION

A few weeks ago, Dr. Gerber telephoned and asked if I would spend a few minutes reminiscing about the past of Frequency Control and the Frequency Control Symposium in this Plenary Session which marks the fortieth anniversary of the Symposium. That request brought the realization that I was indeed one of the "old-timers" in the field, by virtue of having been around for thirty years, although I must confess that it still doesn't seem so very long. I joined Bell Laboratories in 1954, to work in Roger Sykes' Frequency Control Department, and have been involved in various aspects of the frequency control business from that time until the present. Through all that time, the Frequency Control Symposium has remained the unique forum for the field--the one forum at which essentially all the work is reported, and the one forum at which the people involved can assemble to renew old acquaintances and form new ones.

For my part, the career that started so long ago has been a satisfying one. There has never been a shortage of challenging problems in the frequency control field, but we have made a lot of progress. The most enjoyable feature, though, has been the large number of enduring friendships to result from the association with the Frequency Control Symposium, and with the Quartz Devices Section of EIA.

FREQUENCY CONTROL PAST

It seemed appropriate on this occasion to look back at that first symposium to see what kind of work was being done in the field at the time. To put things in proper perspective, a brief statement of the status of Frequency Control at the time is in order.

Most operational frequency and time standards in 1956 were based on quartz crystal oscillators, the most widely used being the 100 kHz GT-cut resonator, with a bridge-type vacuum tube oscillator of the kind developed during World War II as the reference for the LORAN-navigation system. These standards were calibrated in terms of astronomical observations. Frequency aging rates of a few parts in 10^9 /week were achieved, and frequency correlation to about 1×10^{-8} could be realized by means of the standard frequency broadcasts from WWV in the United States. VLF transmissions were being considered for dissemination of standard frequencies.

The possibility of using AT-cut precision, overtone-mode crystal units in Standard Frequency applications was being investigated, and the prototypes of a 2.5 MHz precision oscillator had been delivered to the military. These oscillators, as well as the 5 MHz which was being developed, were also vacuum-tube circuits of the modified-Pierce type.

Silicon Junction transistors had just recently become available, and their application to precision oscillator circuits was one of the areas in which I was involved.

Atomic and molecular transitions as a reference standard of frequency were being investigated, including cesium beam machines and Rubidium gas cells, as well as gas masers.

The transmission method for measuring crystals up to about 30 MHz was being considered as a Standard method, but most crystal measurements were made using the CI-meters. Many problems remained with measurement of overtone crystals, especially at VHF frequencies.

There were many problems recognized in the design of AT-cut crystal units, especially contoured low-frequency units, and also overtone mode VHF crystals.

With that very brief sketch of the status of Frequency Control, it is interesting to note some of the papers which were presented in 1956 at the meeting in the Berkley-Carteret Hotel in Asbury Park. Investigations into the properties of quartz material were reported by Jim King and George Arnold, while work towards optimizing the growth process for synthetic quartz was reported by L. A. Thomas and Danforth Hale.

A mathematical theory of vibrations of plates was reported by R. D. Mindlin, while Van Dyke, Chi, and Hafner all reported observation and measurements of the behavior of AT-cut resonators. Belser and Hicklin reported on aging studies of quartz crystals, and Art Warner presented the design of the 2.5 MHz 5th overtone AT crystal used in a Ground Station Frequency Standard.

John Wolfskill presented a paper describing the problems encountered in the manufacture of precision crystal units, and George Bistline dealt with the manufacture of miniaturized crystal units.

The design of high-frequency crystal filters was the subject of a paper by D. I. Kosowsky.

Professor Dicke reported on the status of atomic and molecular Frequency Standard, and W. D. George presented the performance of a low-temperature frequency standard.

A change-of-state crystal oven, based on use of a substance having a melting temperature at the temperature of operation was described by Smitzer and Strong.

Dr. J. A. Pierce described comparison measurements on Frequency Standards.

Techniques for measurement of the equivalent parameters of VHF crystal units was the subject for papers presented by Guttwein and Pochmerski, and by Robertson.

Judging from this sampling of papers presented at the Tenth Symposium in 1956, it would appear that work was being done in virtually every facet of frequency control.

THE INTERIM

During the ensuing years, the investigations continued, bringing more complete understanding of the properties of synthetic quartz material, which in turn led to improved growing procedures for more repeatable growth of better material. It also led to the recognition of the advantages to be realized by sweeping of synthetic material to remove some of the ions which contribute to anomalous behavior.

Design and processing of quartz crystal resonators continued to improve as better understanding of the vibrational mode structure was gained. New devices were introduced, such as surface acoustic wave resonators and filters, monolithic crystal filters, inverted-mesa VHF fundamental-mode crystal units, and miniature integrated-circuit oscillators.

The atomic and molecular frequency standard work continued leading to the availability of commercial equipment and to the definition of time interval in terms of the Cesium beam frequency. VLF transmissions for the dissemination of standard frequencies appears, along with the quite widespread deployment of LORAN-C transmissions stabilized to cesium-beam equipment--with the results that reference frequency to an accuracy of better than 1×10^{-11} can now be obtained "off the air" essentially anywhere in the USA.

Precision transistor and integrated-circuit oscillators, operating with very precise AT and SC cut crystal units in proportionally-controlled constant-temperature ovens have evolved, to make available miniature, highly stable frequency sources. When calibrated by VLF or LORAN-C signals, these provide excellent working references, available at any development or manufacturing location.

Measurement technology has also advanced significantly, with application of admittance-bridge methods, and scattering-parameter methods, to the precise measurement of crystal unit equivalent parameters. A new measurement standard was ultimately published, about two years ago, defining s-parameter techniques for use up to 1 GHz and beyond, suitable for measurement of both bulk-wave and surface-wave resonators.

Thus, the Frequency Control Symposium has remained the principle forum for presentation of new findings, new ideas, and new device technology in the field of Frequency Control. Much progress has been made during these past three decades toward the generation, definition and dissemination of "ideal" standard frequency and time-interval references, as well as toward the design and manufacture of a wide range of crystal oscillators, crystal filters and crystal resonators for use throughout the telecommunications and data processing industries.

PRESENT AND FUTURE

If now we look briefly at the agenda for the 40th Frequency Control Symposium, our first impression is that things have not changed as much as we thought. Closer examination and comparison to the past reveal that, while problems in the same general areas are still being worked, the level of the problems and the degree of understanding have changed very significantly. For example, in resonator processing, current papers are concerned with the control of triply-rotated cuts, and with refining the etchant baths used for chemical polishing of quartz. For comparison, in the long past, we were concerned with defining the best angles for singly-rotated cuts, and with establishing the "optimum" amount of chemical etching to be carried out in preparation of quartz resonators.

Several papers will be presented here dealing with the design of quartz resonators. Now, however, the power of the computer makes possible the application of finite-element methods, matrix analysis, and nonlinear theory to the design

problem. A paper on the suppression of anharmonic spurious modes would indicate that at least some of the problems recognized thirty years ago have still not been completely solved.

There will also be several papers presented dealing with the properties of quartz material, as well as with Berlinite. The types of impurities and defects in synthetic quartz are much better understood now, as are their influence on the behavior of devices fabricated from the material. However, several questions in this area still remain unanswered.

Work is still going on in the field of atomic and molecular frequency standards, as evidenced by papers dealing with Hydrogen masers, cesium beam apparatus, alkali gas cells, and rubidium equipment. The accuracy and stability of these devices is much improved, but expectations of even further improvement are evident.

The design and analysis of oscillator circuits is seen to occupy several workers, with considerable interest in improved phase noise performance--an area that has been of interest for many years, and some problems still remain.

In summary, it would appear that many exciting challenges remain in the Frequency Control area, despite the progress made during the past three decades. Not only are there unresolved problems in the traditional technologies, but we may expect that advances in electronics, communications, and optoelectronics will result in even closer tolerances on frequency and timing requirements in the future.

THE ATOMICHRON -- 30 YEARS AGO

Richard T. Daly
Chairman and
Chief Executive Officer
Quantronix Corporation
Smithtown, New York

A program to develop a practical, commercial cesium-beam frequency standard was begun in 1954 at National Co., Inc., Malden, Massachusetts. By 1956, prototypes were produced followed by production models. Reminiscences of the two-year effort are presented.

The opportunity to address this 40th anniversary Frequency Control Symposium is a distinct pleasure. I particularly wish to thank Dr. E. A. Gerber whose impressive powers of recall across a gulf of 30 years led to my being invited. Dr. Gerber and his co-workers, mainly Dr. Fritz Reder, played key roles in the development effort which produced the first commercially-available cesium beam frequency standard, the Atomichron. Because of my involvement in that effort, Dr. Gerber has suggested that I provide a few "reminiscences". Unfortunately, my powers of recall are less impressive than his, and to refresh my memory, I have made liberal use of an excellent article by Paul Forman¹ which chronicles much of the effort which was carried out from 1954 to 1956 at National Company, headquartered in Malden, Massachusetts.

The objective of that effort was the stabilization of a relatively low frequency oscillator (5 MHz) against the relatively high-frequency (9 GHz) ground-state hyperfine transition in Cs₁₃₃.

The overall development task at National was actually two subtasks: an electronic one--the design and construction of a system of frequency multipliers and synthesizers together with a servo for locking the master 5 MHz oscillator to the cesium signal, and a beam tube one--the design and construction of an unconventional vacuum tube in which a narrow ribbon beam of neutral cesium atoms issued from a source, passed sequentially through a "state-selector" magnet, a microwave field generated by the multiplier/synthesizer, a "state-analyser" magnet to then impinge on a detector.

The electronic subtask was by no means trivial, given the then-existing state-of-the-art, and its successful execution brought forth several clever designs.

The beam tube task, however, was even more challenging. You will recall that the electronic structure of cesium, like that of the other alkali metals, is characterized by an outer shell with a single, unpaired, spinning electron. This spin gives rise to a magnetic dipole moment. The nucleus, too, has a spin and a magnetic dipole moment. The ground electronic state of cesium, due to the electronic/nuclear magnetic dipole interaction, actually consists, therefore, of two equally populated states whose energies are determined by whether the electronic and nuclear spins are aligned or opposed. Transitions between these two narrowly-separated states provides the stable reference frequency at 9 GHz. Recall, also, that due mainly to the existence of the electronic magnetic dipole moment, the neutral cesium atom will experience a force and can be deflected in a non-uniform (gradient) magnetic field.

In operation, the gradient field of the state-selector magnet separates those atoms issuing from the source in, say, the spins-aligned state, deflecting them through a small angle into the drift region. There, interaction with the magnetic component of the microwave field at or near the resonance frequency reorients ("flips") the relative nuclear/electronic spin configuration to the spins-opposed state. The gradient field of state-analyser then deflects those atoms emerging from the microwave drift region in the spins-opposed state through a small angle onto a detector. The flux of cesium atoms at the detector provides the required servo correction signal.

But, so much for the technology--on to the "reminiscences".

Perhaps the most striking aspect of the Atomichron development was that it was successful at all. It is literally true that, on the day I reported for work as project manager at National's plant, the resources at my disposal consisted of two aging desks, two chairs and one waste basket--humble beginnings, indeed, but better times were to come--the effort soon moved into an empty A&P supermarket in beautiful downtown Melrose, Massachusetts. On inspection, I found that that part of the building to be devoted to vacuum processing had been provided with a newly-laid floor--asphalt black topping!

Looking back from the vantage of 30 subsequent years of managing technically-oriented enterprises, and including the co-founding of my present company, Quantronix, I would have to attribute the success of the Atomichron development largely to the quality of human resources which were ultimately assembled to address the task. The enthusiasm, dedication and aggregate talent represented were truly impressive. But, another important contributing factor was the involvement at the National Company's Board-of-Directors level of Professor Jerrold Zacharias, of MIT, from whose inventive mind sprang the development proposal in the first place.

Interestingly, a third contributing factor was the apparent lack of a "market study". I feel certain, now, that a study, then, would have shown the market to be very small. Conventional business school wisdom would have concluded that National Company did not have the staying power necessary to develop a market and would most certainly have recommended suffocating the infant in its cradle.

But the project did get underway, and, perhaps not unexpectedly, several very basic differences in design approach quickly surfaced. Given the principal players--on the one hand, a freshly-minted Ph.D., myself, armed with "radical" ideas, many of which flew in the face of conventional laboratory molecular beam device configuration, and, on the other, the more conservative Professor Zacharias who had devoted considerable thought to designs which were much more evolutionary--conflicts were inevitable.

One of these differences concerned the length of the beam tube's radio-frequency drift region. It was well-known that the width of the microwave resonance

for a given beam velocity, is inversely proportional to the drift region length. But a long drift region implies a large source-detector separation, which, if the design is restricted to the use of dipole selection magnets, inversely affects the total detected beam flux. The problem bears resemblance to that of an optical system attempting to conserve optical flux using only cylindrical lenses oriented in a common direction.

I argued that a more advanced design should employ multipole magnets for axially-symmetric deflection (spherical lenses) in which case the length of the drift region would play a substantially-lesser role in determining signal strength. In short, it was not at all clear that with dipole magnets, the signal-to-noise and thus the resolution of the resonance was optimized with a long drift region. Later developments at other organizations have confirmed the correctness of the short-tube, multipole magnet concept, but the compromise which was struck at National committed us to a 6-foot beam tube.

Having made the decision to go with a "long" beam tube, the question of the beam tube envelope loomed large. Clearly, in the interest of reliability, not to mention cost, an initially-evacuated and permanently-sealed tube could be seen as an important objective. Just as clearly, a glass envelope with everything save the source and detector external to the tube was the way to go. But, if the design called for a 6-foot long tube to be fitted into magnet gaps of 1/4" or so, the unacceptable fragility of a glass structure was overriding and the use of a non-magnetic metal, 300 series stainless steel, was required. This compromise was adopted. However, it led to further problems, chief among them being vacuum integrity. Fortunately, a successful band-aid solution, in the form of an ion-getter pump, proposed by Zacharias himself, proved successful. Since it was non-mechanical, no vibration was introduced and, except for long periods of non-operation, it was able to handle the relatively slow outgassing of the metal beam tube.

After a two-year effort, a very credible prototype of the Atomichron emerged and was introduced in October 1956 at the Overseas Press Club in New York City with great fanfare. Five more "pre-production" units were then in the pipeline, and, over the next 5 years, 50 units were produced. All exceeded the original specification by about an order of magnitude.

At the time of the introduction of the prototype in New York, I had concluded that National was in no shape to continue an intensive development program and shortly thereafter, I left.

Notwithstanding the occasional "clashes of will" and the fact that not one, but several versions of the wheel were re-invented during the development, the fact of the program's ultimate technical success calls for comment.

It is my view that even if a larger organization with an in-place high-technology development staff and facilities had been willing to undertake the Atomichron project, success would have been less assured. I lay this to two highly-motivating factors at National Company--a "carrot" and a "stick". At National there were only two levels of management: top and middle. The direct and frequent communication between these levels made the "stick" very obvious. This was a "sink-or-swim, bet the Company"

project--it just had to play. The "carrot" part was also very much in evidence: it was just enormous fun re-inventing all those versions of the wheel.

¹Paul Forman, "Atomichron: The Atomic Clock from Concept to Commercial Product", Proc. IEEE, Vol. 73, p. 1181, July, 1985.

COMPUTER MODELING OF POINT DEFECTS IN QUARTZ

T. M. Wilson, L. E. Halliburton, M. G. Jani and J. J. Martin,
Department of Physics
Oklahoma State University
Stillwater, Oklahoma 74078-0444.

Abstract

A wide spectrum of experimental techniques are used to investigate point defects in quartz. Among these are infrared absorption, acoustic loss, dielectric loss, electron spin resonance (ESR), and electrical conductivity. As data from these experiments accumulate, it is becoming apparent that further progress in the study of point defects will require a direct interaction with theory. The complexities of quartz (e.g., non-cubic symmetry, large c-axis channels, and mixed ionic-covalent bonding) cause many experimental results to be inconclusive with regard to defining models and identifying basic mechanisms. Thus, theoretical modeling of a defect according to the dictates of quantum mechanics will be a direct aid in the interpretation of experimental data.

We describe a general approach to the computer modeling of point defects in quartz. The method utilizes recently developed quantum chemistry computer programs which perform *ab-initio*, self-consistent-field, molecular orbital calculations on large clusters of atoms. Thus, we model a macroscopic crystal of quartz by molecular clusters, where the atoms are positioned relative to each other as they would be in the perfect lattice. This scheme easily provides for placing a defect such as an interstitial alkali ion, a substitutional aluminum ion, or an oxygen vacancy near the center of the cluster. In general, the computer programs determine the total energy of the cluster. By varying the positions of the atoms within and surrounding the defect, a minimum in the cluster energy is found which corresponds to the defect's equilibrium configuration. The defect models that emerge from these calculations have proved useful in interpreting acoustic-loss measurements on the aluminum-lithium, Al-Li, and aluminum-sodium, Al-Na, centers and ESR results from interstitial-lithium-associated electron traps. They show the sodium in the Al-Na center to lie off the x axis with minima in the (±y, ±z) positions; the lithium in the Al-Li center to lie on the x axis with a single minimum; and the lithium in the interstitial-lithium-electron trap to lie on the twofold x axis near the center of the c-axis forming a stable S=1/2 defect.

Introduction

Defects play an important role in many of the applications of quartz.¹⁻⁴ For example, the Q value of a quartz resonator is known to directly depend on impurities introduced during the growth of the crystal. Additional defect-related problems arise when ionizing radiation modifies pre-existing growth defects and creates new defects. The large variety of interesting defects, formed either during growth or by radiation, arises from several of the unique features of quartz (e.g., the non-cubic symmetry, the presence of large c-axis channels, and the mixed ionic-covalent bonding).

Point defects in quartz continue therefore to be the subject of intense study involving a wide spectrum of experimental techniques, ranging from optical to magnetic resonance to transport measurements. Furthermore, it is becoming widely recognized that investigations of point defects have reached the point where further progress requires direct interaction with theory to aid in the interpretation of experimental data. With the development of high-speed computers, theoretical activity has picked up significantly.

Here we describe a general approach to the computer modeling of point defects in quartz. The method utilizes recently developed, highly sophisticated, quantum chemistry computer programs which perform *ab-initio*, self-consistent-field, molecular orbital calculations on large clusters of atoms.^{5,6} From these calculations we obtain: total and one-electron energies of the defect as a function of the positions of the cluster atoms, charge densities and net ionic charges, spin densities and hyperfine parameters, information on the nature of the bonding, vibrational frequencies and their polarization, and the optimized positions of the atoms surrounding the defect. The results of these calculations, combined with the experimental measurements, are providing important, new insight into the electronic and vibrational properties, as well as atomic scale models, of point defects in quartz.

One begins by simulating a macroscopic crystal of quartz by a molecular cluster such as that shown in Fig.1. Here the atoms are positioned relative to each other as they would be in the perfect quartz lattice. Hydrogen atoms are used to terminate the cluster in order to eliminate problems associated with unpaired oxygen electrons at the edge of the cluster. The hydrogens are positioned along the Si-O bond directions at a distance of 0.95 Å from the oxygen, (the OH bond distance). This scheme easily allows for the placing of a defect such as an interstitial ion, a substitutional aluminum ion, or an oxygen vacancy near the center of the cluster. The particular choice of model cluster depends upon the defect one intends to study. In general, the programs determine the minimum total energy of the cluster. By varying the positions of the atoms surrounding the defect, the cluster's total energy is minimized. This minimum corresponds to the defect's equilibrium configuration. Throughout the minimization process, the positions of the hydrogens remain fixed, thereby simulating the restraining effect of the quartz lattice.

To illustrate the utility of this approach, we shall describe the defect models that have emerged from these calculations for the Al-Li and Al-Na centers, and for the interstitial-lithium-associated electron trap. These models are proving quite useful in the interpretation of acoustic- and dielectric-loss measurements on the Al-Na and Al-Li centers, and ESR and thermally stimulated luminescence (TSL) measurements of the interstitial-lithium-associated electron trap center in quartz.

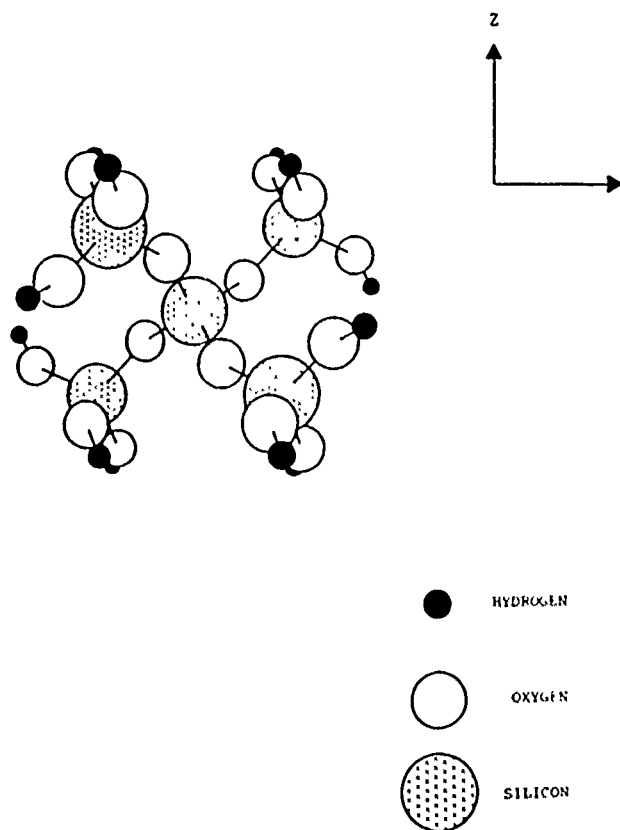


Fig. 1. Thirty three atom cluster model of α -Quartz.

Aluminum-Alkali Centers

As-grown cultured quartz contains substitutional aluminum which is charge compensated by interstitial alkalis. If an as-received crystal is exposed to ionizing radiation at temperatures above 200 K, the aluminum-alkali centers are converted into a mixture of Al-OH and Al-Hole centers.⁷⁻⁹ Schematic representations of these aluminum-related centers are given in Fig. 2. The aluminum-alkali centers exhibit no characteristic infrared bands and, being diamagnetic, are unobservable using ESR techniques. However, they do cause a local lattice deformation whereby they can couple to external forces acting along certain crystallographic directions, and can undergo a deformation relaxation with an attendant mechanical loss.

In contrast to earlier work¹² which reported a Li-related loss peak near 105 K, the results of the acoustic loss measurements made by Martin¹⁰ for the Premium Q blanks and similar results for the Toyo samples show no evidence of an Al-Li acoustic loss at temperatures below 100 °C. The absence of Li-related dielectric loss peaks in both Sawyer Premium Q and Toyo Supreme Q samples have also been recently reported.^{13,14} From these results a model has been proposed^{13,14} where, due to the small size of the Li^+ ion, the double well associated with the Na has have collapsed into a single well for the Li ion. The resulting Al-Li pair would then lie along the twofold x axis. Such a single well model would show neither acoustic- nor dielectric-loss peaks.

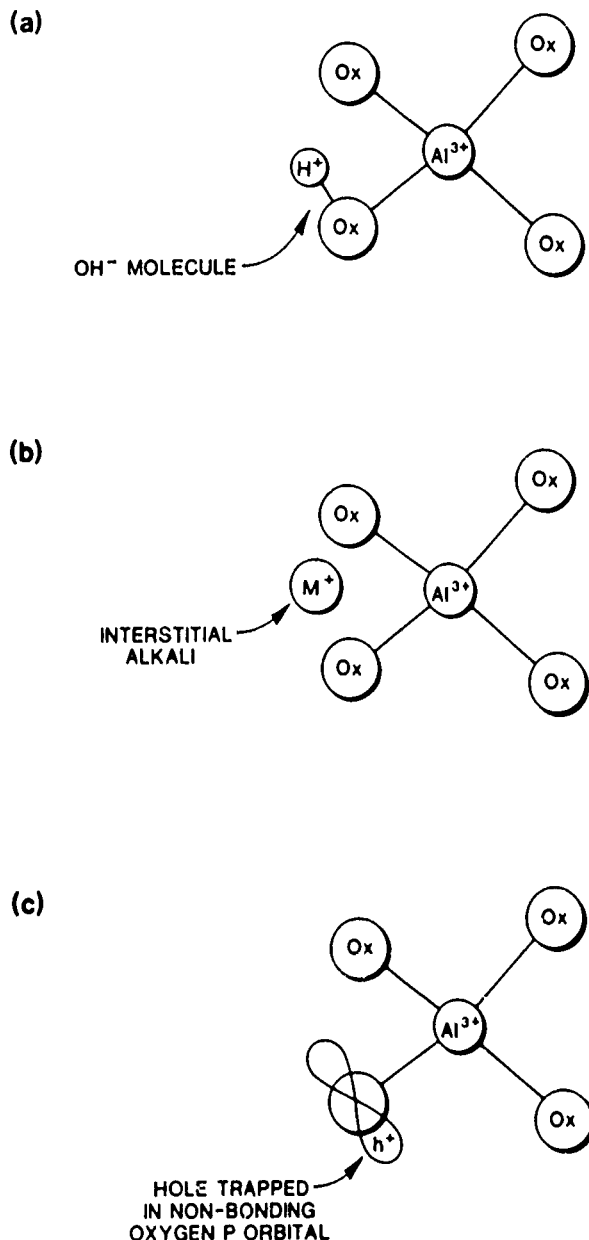


Fig. 2. Schematic representation of the three primary aluminum-associated defect centers in quartz, (a) the Al-OH center, (b) the Al- M^+ center with M^+ either Li^+ or Na^+ , and (c) the Al-Hole center.

Figure 3 compares the acoustic loss Q^{-1} spectra for unswept, Na-swept, and Li-swept resonator blanks, all of which were fabricated from the same bar of Sawyer Premium Q quartz.¹⁰ The unswept blank shows a small Al-Na loss peak at 53 K. Lithium sweeping removes this peak and introduces no new peaks. The Na-swept blank shows a very large 53 K peak. A much smaller loss peak related to the Al-Na center is observed at approximately 135 K. Park and Nowick¹¹ have also observed two Na-related peaks in their dielectric loss measurements.

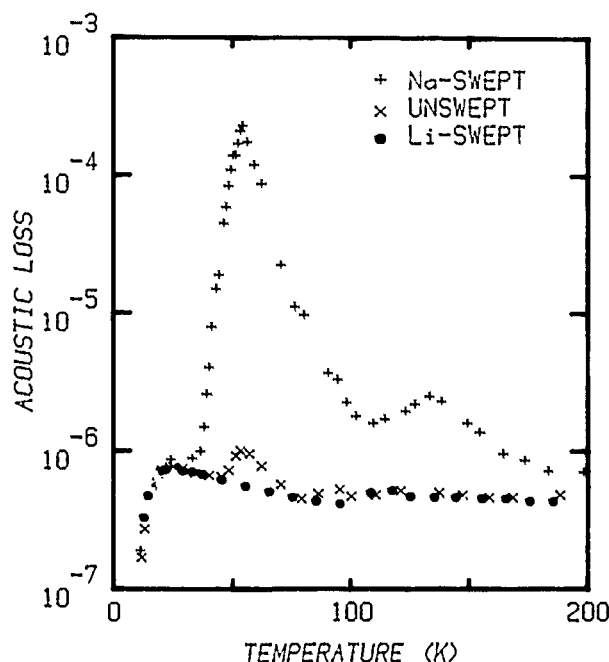


Fig.3. Acoustic loss spectra of unswept, Na-swept, and Li-swept resonator blanks all fabricated from the same bar of Sawyer Premium Q quartz are shown. The Al-Na center is responsible for the peak at 53 K.

The two models employed in our studies of the aluminum-alkali centers are based upon those shown in Fig. 4. Both clusters display C_2 symmetry with respect to rotation about the x axis. For the calculations which utilized the nine-atom cluster shown at the top of Fig. 4, the silicon was replaced by an aluminum and an alkali atom was placed at a distance of 2.5\AA from the aluminum along the x axis. The calculations employing the 18-atom cluster shown at the bottom of Fig. 4 take into account the influence of the SiO_4 group on the far side of the c axis on the position of the alkali. In this case, the aluminum-alkali center was modeled by replacing one of the silicon atoms by an aluminum, and an alkali was introduced midway along the x axis connecting the Si and Al atoms.

Geometry optimization calculations have been carried out for the ten-atom cluster models of the Al-Li and Al-Na centers. Our results agree closely with those recently published by Mombourquette and Weil.^{15,16} They show that the lithium lies within 0.001\AA of the x axis and has a single minimum. The sodium is displaced off the x axis by $\sim 0.3\text{\AA}$ in the $(-z, -y)$ direction. There is a second minimum in the $(+z, +y)$ direction. Their results support the model proposed by Toulouse et al.^{13,14} to account for the absence of either a dielectric- or acoustic-loss peak associated with the Al-Li center.

The ten-atom cluster models for the Al-Li and Al-Na centers neglect the interactions of the alkali with the SiO_4 group on the opposite side of the c-axis channel. Since the alkali-oxygen and alkali-silicon distances are roughly the same as those for the AlO_4 group atoms, we carried out calculations using the

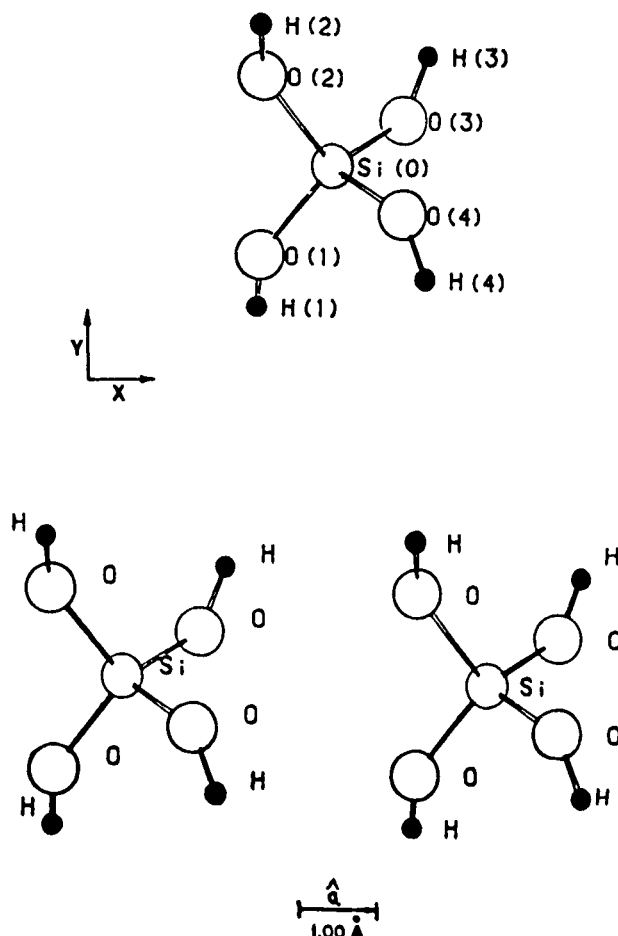


Fig.4. Nine-atom and eighteen-atom primary models used in the SCF-MO calculations described in this paper.

more realistic nineteen-atom cluster models shown in Figs. 5 and 6 for the Al-Na and Al-Li centers, respectively. In these calculations, the alkali was initially placed midway between the aluminum and silicon atoms along the x axis. The AlO_4 group atoms were placed at the displaced positions predicted by the ten-atom cluster calculations. The dotted circles represent the positions the atoms would occupy in the undistorted quartz lattice. Geometry optimizations were made where (i) the alkali was allowed to move while holding the rest of the atoms in the cluster fixed and (ii) the alkali and the four atoms pointing into the c-axis channel were allowed to move while holding the rest of the atoms fixed. These calculations gave results that were in qualitative agreement with those obtained using the ten-atom clusters. The sodium is displaced off the x axis with minima in $(\pm y, \pm z)$ positions although the displacement is roughly half that predicted by the ten-atom cluster results. The lithium remained on the x axis, slightly closer to the aluminum than to the silicon. Calculations were made where the lithium was moved off the x axis to one of the positions where the sodium has an energy minimum. Geometry optimization resulted in the lithium moving back onto the x axis, and to its having a single energy minimum. Although our results for the nineteen atom cluster are not yet fully optimized, they clearly show that the SiO_4 group on the opposite side of the c-axis channel significantly influences the final equilibrium position of the alkali.

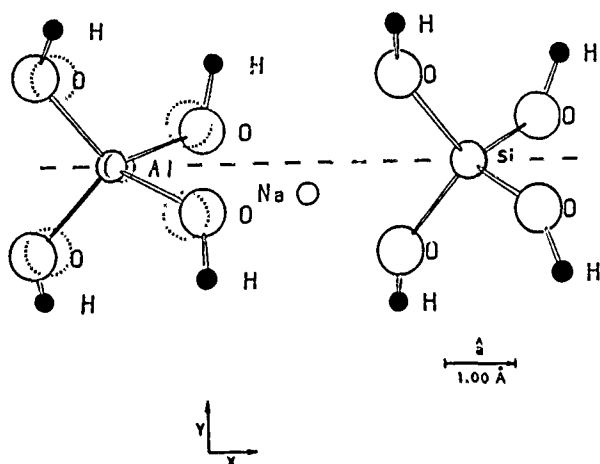


Fig.5. Nineteen-atom model of the Al-Na center. Dotted circles indicate the positions of the atoms in the undistorted quartz lattice.

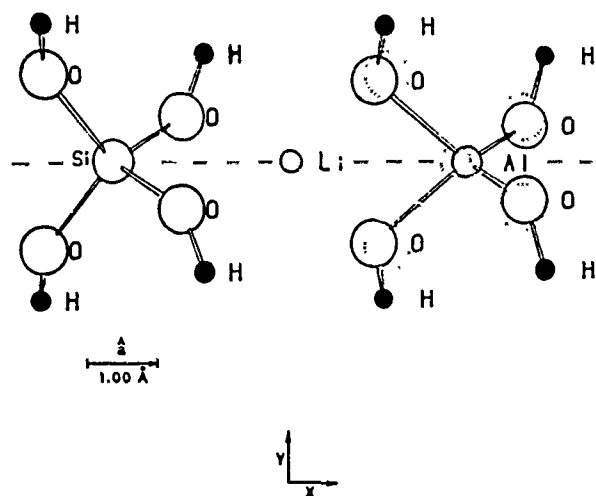


Fig.6. Nineteen-atom model of the Al-Li center.

The results of the model calculations described above for the Al-Na and Al-Li centers support the models previously suggested^{13,16} based upon the results of acoustic-loss and dielectric-loss measurements. The calculations also reveal other important differences between the sodium and lithium atoms in the aluminum-alkali centers. They show that the sodium atom gives up one of its electrons to the AlO_4 group and serves as a charge compensator in much the same fashion as previous models for this center have suggested. However, the lithium atom in the Al-Li center behaves much differently, and contrary to

previously held ideas about the behavior of alkalis in quartz. Situated as it is, displaced slightly off the center of the c-axis towards the aluminum along the x axis, the nearest neighbor Li-O distances are ~ 1.9 Å. This distance is nearly ideal for the formation of Li-O molecular orbitals, whereas it is too short for the formation of stable Na-O molecular orbitals. An examination of the cluster's molecular orbitals reveals that the sp^3 hybrid lithium orbitals form covalent bonds with the four nearest neighbor oxygens and has a net ionic charge of \sim zero, whereas the sodium in the Al-Na center retains its ionic character. This result provides the reason why the lithium remains on the x axis as it requires too much energy to distort the Li-O covalent bonds. It may also account for the difference in the behavior of lithium in quartz compared to that in other, predominantly ionic crystals, where substitutional lithium is known to favor an off-center position.¹⁷

Lithium-Associated $S=1/2$ Electron Trap

It is well known that by exposing a quartz crystal to ionizing radiation above 200 K, the aluminum-alkali centers are converted into a mixture of Al-OH and Al-hole centers,⁷⁻⁹ shown schematically in Fig.2. The liberated alkalis move along the large c-axis channels and become trapped at an as yet undetermined site in the quartz lattice. Naively, one might expect that the interstitial alkali could trap an electron during irradiation at low temperatures and form an alkali atom analogous to the hydrogen atom found in quartz.^{18,19} Jani, Halliburton and Halperin²⁰ recently reported on their discovery of a simple lithium-associated electron trap in quartz. The defect consists of an extra electron trapped at a four-fold-coordinated silicon atom that is stabilized by an adjacent interstitial lithium. The defect is formed by a double irradiation sequence, first between 150-300 K to release the lithium ion from its associated aluminum, then at 77 K to trap an electron at the interstitial lithium.

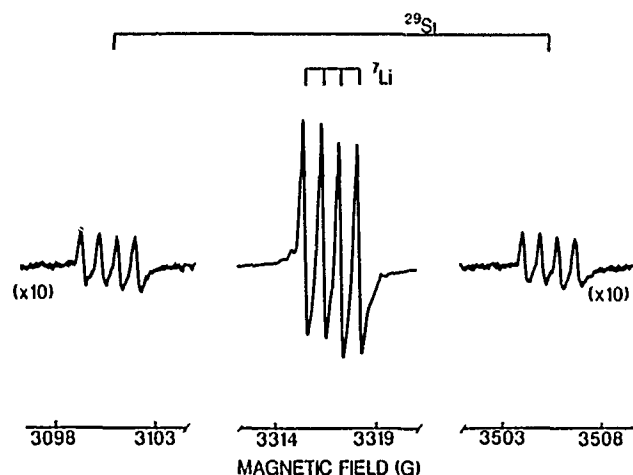


Fig.7. ESR spectrum of the $[\text{SiO}_4/\text{Li}]^0$ center. These data were taken at 77 K with the magnetic field parallel to the crystal's c-axis. The gain was increased by a factor of ten when recording the two outer sets of four lines.

The c-axis ESR spectrum shown in Fig.7, displays a hyperfine splitting of 0.9 G from a ${}^7\text{Li}$ and 404.7 G from a ${}^{29}\text{Si}$. The defect clearly does not resemble that of a lithium atom. The analysis of the angular dependence of the primary set of ESR lines arising from defects with no ${}^{29}\text{Si}$ nucleus shows that the interstitial lithium lies on a twofold axis passing through the adjacent silicon having the large spin density. Following the notation scheme proposed by Weil,³ they labeled this new defect the $[\text{SiO}_4/\text{Li}]^0$ center.

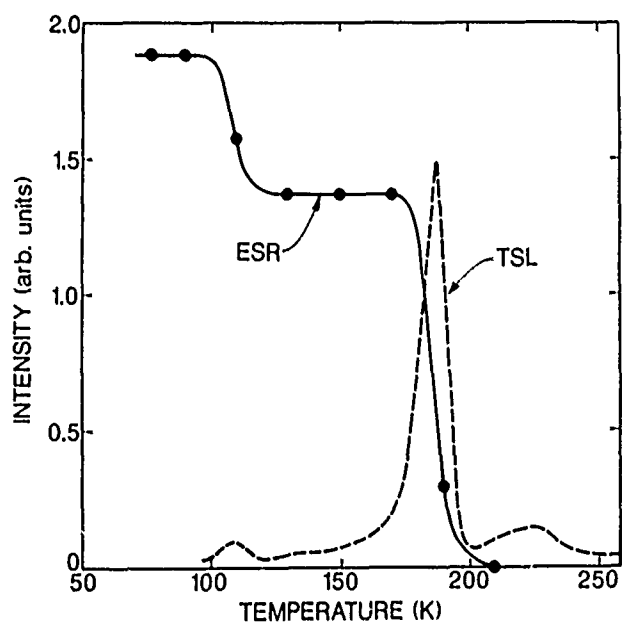


Fig.8. Thermal anneal behavior of the $[\text{SiO}_4/\text{Li}]^0$ center's ESR spectrum. A glow curve containing the 190 K peak is plotted on the same temperature scale for comparison.

The results of a pulse-anneal experiment to determine the stability of the center is shown in Fig.8. Each data point was taken at 77 K and represents the amount of ESR signal remaining after the sample was held at the indicated anneal temperature for 5 minutes. The defect has two decay steps, one near 109 K and the other near 187 K. Jani et al.²⁰ suggest that the 109 K step occurs when holes are released from other traps in the crystal and migrate to the $[\text{SiO}_4/\text{Li}]^0$ centers. The 187 K step corresponds to the intrinsic decay of the center. This interpretation is further reinforced by the appearance of a thermally stimulated luminescence peak near 190 K,²¹ which is also shown in Fig.8. Their analysis of the kinetics of the TSL peak gives a thermal activation energy of 0.6 eV, and suggests that the unpaired-spin electron's energy level should lie below the bottom of the conduction band.

Model calculations were made on this new defect where the nineteen-atom neutral cluster shown in Fig.9 was used to simulate the $[\text{SiO}_4/\text{Li}]^0$ center.²² The two SiO_4 groups in this cluster are situated on either side of the c-axis channel in the quartz lattice with a lithium lying along the twofold symmetry axis connecting the two silicons. All the atoms, with the exception of the hydrogens, were allowed to move

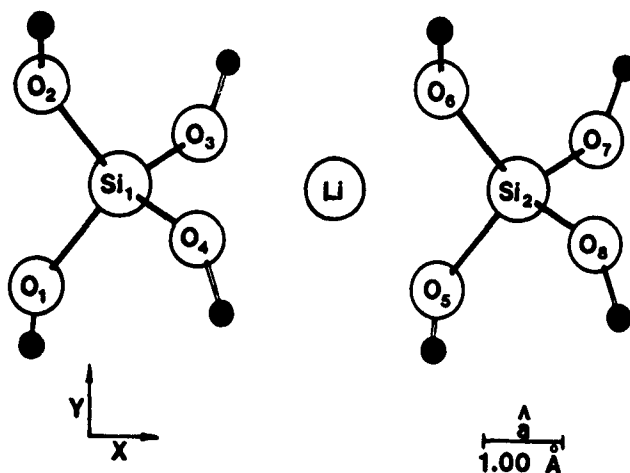


Fig.9. Nineteen-atom model of the $[\text{SiO}_4/\text{Li}]^0$ center.

during the geometry optimizations. No conditions forcing twofold symmetry were imposed. The final equilibrium positions obtained from this procedure show the lithium and the two silicons to lie within 0.001 Å of the x axis with the lithium atom positioned slightly closer to Si_2 than to Si_1 (by 0.05 Å). The calculated spin density at the silicon nucleus Si_2 to that at the lithium is ~61 and to that at Si_1 is ~217. The values of the experimental silicon and lithium hyperfine constants suggests that the value for $\rho_{\text{Si}}/\rho_{\text{Li}}$ for the silicon nearest the lithium should be on the order of 10^2 , and the spin density on the second silicon is negligibly small. A similar asymmetry in the two silicon spin densities has been observed for the interstitial silver atom, which also lies in the c-axis channel along the twofold axis connecting the two silicon ions.²³

A surprising and important result to come out of these calculations was the prediction of a net ionic charge for the lithium of approximately zero ($Q_{\text{Li}} \sim -0.05 e$), yet at the same time predicting spin densities at the ${}^7\text{Li}$ and ${}^{29}\text{Si}$ nuclei that agree quite well with the measured values. An analysis of the molecular orbitals shows that the lithium forms sp^3 hybrid orbitals and participates strongly in the formation of covalent bonds with the four oxygens that point in towards the c-axis channel and also forms sp_x bonds with the two silicons. This results in the spin density due to the trapped, unpaired-spin electron being shared by the lithium and the SiO_4 group centered on Si_2 . The calculations further show the position of the unpaired-spin electron's energy level to be ~0.8 eV below the bottom of the conduction band, in good agreement with the experimentally determined thermal activation energy.²¹

Summary

Most aspects of the results of the model calculations described here for the aluminum-alkali and $[\text{SiO}_4/\text{Li}]^0$ centers in quartz agree very well with the results of EPR, TSL, acoustic- and dielectric-loss experiments. In addition, the results obtained from these calculations have provided new and important insights into the microscopic nature of these defects.

Clearly, the results obtained from these model calculations are limited in accuracy by many considerations. Among these are the modest size of the clusters used, the fixed location and electronic nature of the terminating protons, the need for using modest-sized basis sets, and the neglect of configuration interaction. However, the results presented here, as well as the work of J. A. Weil and his group at Saskatchewan, and of W. B. Fowler and his group at Lehigh, suggest quite strongly that the short range effects are most important in determining the ground state properties of defects in quartz. It is now clear that the results of these types of calculations, when combined with experimental measurements, will play an increasingly important role in determining the electronic properties, as well as atomic scale models, of point defects in quartz.

References

1. D. B. Fraser, in Physical Acoustics, edited by W. P. Mason (Academic Press, New York, 1968), Vol. V, pp. 59-110.
2. E. J. Friebale and D. L. Griscom, Treatise on Materials Science and Technology, edited by M. Tomozawa and R. H. Doremus (Academic Press, New York, 1979), Vol. 17, pp. 257-351.
3. J. A. Weil, Phys. Chem. Minerals **10**, 149 (1984).
4. L. E. Halliburton, Cryst. Latt. Def. and Amorph. Mat. **12**, 163 (1985).
5. J. S. Binkley, R. A. Whiteside, R. Krishnan, H. B. Schlegel, R. Seeger, D. J. DeFrees, and J. A. Pople, Quantum Chemistry Program Exchange # 406, Indiana University, Bloomington IN.
6. J. S. Binkley, M. Frisvh, K. Raghavachari, D. DeFrees, H. B. Schlegel, and J. A. Pople, Gaussian 82, Release A, Carnegie-Mellon University, Pittsburgh, PA.
7. M. E. Markes and L. E. Halliburton, J. Appl. Phys. **50**, 8172 (1979).
8. W. A. Sibley, J. J. Martin, M. C. Wintersgill, and J. D. Brown, J. Appl. Phys. **50**, 5449 (1979).
9. L. E. Halliburton, N. Koumvakalis, M. E. Markes, and J. J. Martin, J. Appl. Phys. **52**, 3565 (1981).
10. J. J. Martin, J. Appl. Phys. **56**, 2536 (1984).
11. S. Park and A. S. Nowick, Phys. Status Solidi A **26**, 617 (1979).
12. D. B. Fraser, J. Appl. Phys. **35**, 2913 (1964).
13. J. Toulouse, E. R. Green, and A. S. Nowick, Proc. 37th Annual Symposium on Frequency Control (1983), pp. 125-129.
14. J. Toulouse and A. S. Nowick, J. Phys. Chem. Solids, **46**, 1285 (1985).
15. M. J. Mombourquette and J. A. Weil, Can. J. Phys. **63**, 1282 (1985).
16. M. J. Mombourquette and J. A. Weil, J. Magn. Res. **66**, 105 (1986).
17. A. S. Barker, Jr. and A. J. Sievers, Rev. Mod. Phys. **47**, Suppl. No. 2 (1975).
18. B. D. Perlson and A. J. Weil, J. Magn. Res. **15**, 594 (1974).
19. J. Isoya, J. A. Weil and P. H. Davis, J. Phys. Chem. Solids **44**, 335 (1983).
20. M. G. Jani, L. E. Halliburton, and A. Halperin, Phys. Rev. Lett. **56**, 1392 (1986).
21. A. Halperin, M. G. Jani, and L. E. Halliburton, Phys. Rev. B (in press).
22. T. M. Wilson, J. A. Weil, and P. S. Rao, Phys. Rev. (to be published).
23. P. H. Davis and J. A. Weil, J. Phys. Chem. Solids **39**, 775 (1978).

A RADIATION GROWTH STUDY OF ACOUSTIC LOSS
RELATED DEFECTS IN ALPHA-QUARTZJ. J. Martin, Ho B. Hwang, and T. M. Wilson
Department of Physics
Oklahoma State University
Stillwater, OK 74078-0444.ABSTRACT

As-grown cultured quartz contains substitutional aluminum which is charge compensated by interstitial alkalis. Protons trapped at as-yet unidentified sites form OH⁻ related growth-defects which are responsible for the infrared bands observed at low temperatures. Of the above defects, only the Al-Na center with a strong acoustic loss peak at 53 K and a weak peak at 135 K produce observable acoustic loss below 400 K in as-grown quartz. If the crystal is exposed to ionizing radiation at temperatures above 200 K the aluminum-alkali center is converted into a mixture of Al-OH⁻ and Al-hole centers. Acoustic loss peaks at 23 K, 100 K and 135 K are observed to grow in with increasing radiation dose. In Na-swept low defect content quartz the Al-Na center follows a single exponential decay with radiation dose. In quartz with a higher aluminum content the center seems to decay with a double exponential dose dependence. A positive frequency shift upon irradiation which tracked with the aluminum content was observed for the Na-swept samples. This shift was caused by the removal of the 53 K Al-Na acoustic loss peak. Because of a change in the temperature dependence of the background frequency the shift at the turn-over temperature was smaller than expected. Irradiation of a Li-swept BT-cut crystal produced both the 23 K and 100 K loss peaks. The 100 K Al-hole center peak was larger than the 23 K peak; this result is the reverse of our observations in AT-cut crystals. The introduction of these two peaks into the loss spectrum caused the expected negative frequency shift at room temperature. A subsequent irradiation at 80 K of the Li-swept BT-cut sample enhanced both the 23 K and 100 K loss peaks. An initial irradiation at 80 K of a H-swept AT-cut crystal produced both the 23 K and 100 K peaks. Here the peaks had a nearly equal strength; our past observations of room temperature irradiated AT-cut samples showed that the 23 K peak was 3 to 4 times larger than the 100 K peak. The production of the 100 K peak by the initial low temperature irradiation of the H-swept crystal and its enhancement in the Li-swept crystal are consistent with the assignment of the peak to the Al-hole center.

INTRODUCTION

Because quartz crystal controlled oscillators are used in an increasing number of applications where radiation effects must be considered there is increased interest in radiation effect studies¹⁻⁴. The crystal may exhibit both steady-state and transient frequency shifts and Q or resistance changes when exposed to ionizing radiation⁵⁻⁸. Both positive and negative steady state frequency shifts are observed. The early work of King⁹ and other investigators¹⁰⁻¹³ showed that these effects were often associated with the presence of point defects. The best understood of these defects is the substitutional aluminum impurity. In as-grown synthetic quartz the aluminum is charge compensated by an interstitial alkali ion; in high quality material such as Sawyer Premium Q quartz lithium is the dominant alkali¹⁴. When irradiated at temperatures above about 200 K the alkali ion moves away and is replaced by either a proton or a hole located on an adjacent oxygen. Thus, the irradiation forms a mix of Al-OH¹⁵ and Al-hole¹⁶ centers at the aluminum site. The alkali is subsequently trapped at

an as yet unknown site; the proton most likely was released from one of the as-grown OH defect sites. A recent study by Norton, Cloeren, and Suter² suggests that at low radiation levels the aluminum content does not directly affect the radiation response. Other defects such as substitutional germanium¹⁷ and a variety of oxygen vacancy centers also play a role in the radiation process¹⁸. Those centers that directly affect the frequency of a crystal are expected to show acoustic loss peaks.

Fraser¹⁹ has reviewed anelasticity in quartz. The contribution of defect such as the Al-Na center to the acoustic loss of a crystal can be written as

$$Q^{-1} = 2(Q^{-1})_{\max} \omega \tau / (1 + \omega^2 \tau^2). \quad (1)$$

Here Q^{-1} is the contribution which is superimposed upon the normal thermal background, $(Q^{-1})_{\max}$ is the height of the peak, ω is the angular frequency and τ is the relaxation time. The relaxation time is usually taken to be $\tau = \tau_0 \exp(E/kT)$ where τ_0 contains the "jump time" and an entropy factor. The process assumes that the defect is thermally relaxing or jumping between two, or more, energetically equivalent orientations separated by a barrier height E . The crystal vibration at frequency ω then modulates the depth of the wells on either side of the barrier. The maximum loss takes place at $\omega \tau = 1$. This defect related acoustic loss is, of course, the direct analog of the well known dielectric loss. Unlike dielectric loss, acoustic loss measurements are restricted to the normal modes of vibration of the crystal under study. Thus, these measurements are usually made at fixed frequency with the vibration brought into "resonance" with the defect by changing the temperature to get $\omega \tau = 1$. At this temperature the loss is maximum and we often describe the loss peak by giving its peak temperature. For example, the Al-Na center has a well known strong peak at 53 K for 5 MHz crystals. Since the loss due to the defect is small we can write the frequency, f , of the fundamental AT shear mode as

$$(f - f_b)/f_b = (Q^{-1})_{\max} / (1 + \omega^2 \tau^2). \quad (2)$$

f_b is the temperature dependent frequency of the crystal in the absence of the defect. At low temperatures where $\omega \tau \gg 1$, f is greater than f_b and at high temperatures where $\omega \tau \ll 1$, it approaches f_b . Thus, the maximum frequency shift when the sample goes through a loss peak is $(Q^{-1})_{\max}$. This defect related variation in frequency with temperature is superimposed upon the general temperature dependence due to the elastic constants which is contained in f_b . The Al-Na center has a strong loss peak at 53 K when the crystal is irradiated the sodium is removed from the defect and the loss is greatly reduced. Thus, at the normal operating temperature of the crystal we would expect a positive frequency shift equal to the change in the height of the loss peak. If irradiation produced new loss peaks at temperatures below the operating temperature of the crystal they should cause negative shifts. The total shift is expected to be the sum of the changes in the heights of the loss peaks.

At the present time, the Al-Na related loss peaks at 53 K and 135 K are the only ones that we have

observed in as-grown synthetic quartz. Irradiation at room temperature produces peaks at 23 K, 100 K and 135 K. King as cited in King and Sander⁵ first observed the 100 K peak following a low temperature irradiation and associated it with the Al-hole center. Jones and Brown²⁰ reported the production of a loss peak near 20 K by ionizing radiation that matches the 23 K peak. Martin¹⁴ showed that these three peaks anneal out at the same temperature as the Al-hole center. However, the 23 K peak production characteristics seem to be somewhat different than those of the Al-hole center¹⁶. Koehler and Martin⁸ also observed alkali related peaks at 340 K (Na) and 305 K (Li) which grew rapidly with low doses and then decayed for higher doses. Martin, Hwang and Bahadur²¹ have observed that the 340 K Na-associated peak is present in at least some as-swept material; the Li-associated peak at 305 K may also be present but it seems to be weaker than the Na peak. The 53 K and 135 K peaks present in Al-Na center containing quartz are the only loss peaks that have been positively associated with a specific defect. While the peaks at 340 K and 305 K are definitely associated with the presence of Na and Li respectively the actual trap site remains unknown. The radiation produced peak at 100 K is probably related to the Al-hole center but the identity of defects responsible for the peaks at 23 K and 135 K remains uncertain. The main objectives of this project are to investigate the production by ionizing radiation of the defects responsible for acoustic loss peaks in high quality synthetic quartz.

EXPERIMENTAL PROCEDURE

Most of the samples used in this study were 5 MHz AT-cut overtone crystals cut from pure-Z growth lumbered bars of synthetic quartz. Two sets of Warner design 5th overtone blanks have been fabricated from Sawyer Premium Q material; one set taken from bar PQ-E which has 10-15 ppm aluminum and the second from a bar designated SP-DD which has less than 1 ppm aluminum. The bar designated PQ-E was from Sawyer autoclave run K19-5 while SP-DD was from run D14-45. The PQ-E blanks were fabricated by K-W Mfg., and the SP-DD blanks by Piezo Crystal Co. A third set were 3rd overtone AT- and BT-cut blanks fabricated by FEI from our Toyo Supreme Q bar SQ-B which also has 10-15 ppm aluminum. Lithium, sodium, and hydrogen sweeping runs were then made on the originally unswept blanks. "Finished crystals" were made by vapor depositing 5.1 mm diameter gold electrodes and cementing the electroded blank in a "TO" style holder with Epotek P-1011 polyimide. The cover was not installed on the holder. All measurements were made with the crystal mounted in an evacuated cryostat.

We use either the log-decrement method or a transmission method to measure the acoustic loss versus temperature spectrum. Since the transmission method can be automated as described below it is our preferred technique. However, the log-decrement method works better for crystals with large ($>10^{-4}$) Al-Na loss peaks. The acoustic loss spectrum was measured over the 9 K to 330 K temperature range with the crystal mounted on the cold stage of a CTI closed-cycle helium refrigerator. The Q^{-1} was calculated from the crystal series resistance as measured by the transmission system shown in Fig. 1. The frequency output of the synthesizer is controlled by a digital feedback loop which keeps the system tuned to "zero-phase". The loop uses the analog phase output from the vector voltmeter as the error signal which is sensed by the DMM and then sent over the IEEE-488 bus to the computer which controls the synthesizer. The computer also keeps track of the temperature by periodically interrogating the temperature controller. When the temperature has changed the desired amount (usually 2 or 3 K) the relay is

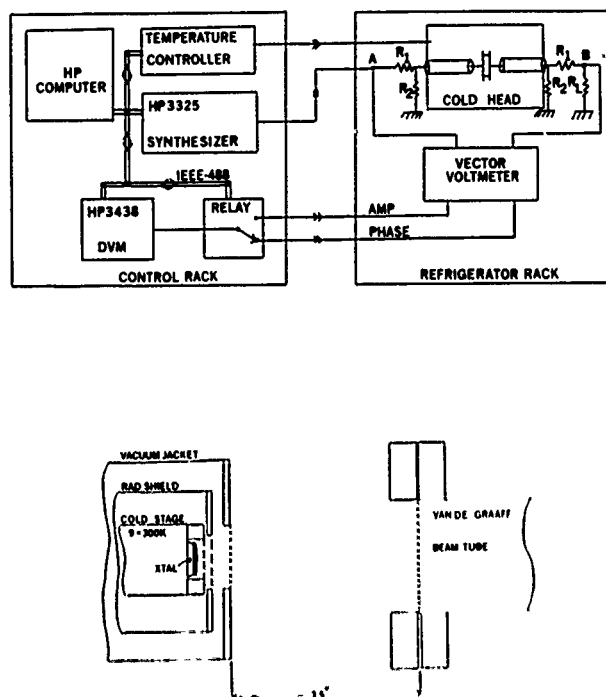


Fig. 1. The block diagram of the transmission system for crystal resistance measurements is shown. The lower portion shows the electron irradiation geometry.

switched so that the DMM reads the amplitude output corresponding to V_g . The computer then calculates the resistance, shifts the relay back to the phase signal and continues holding the frequency on track. The resistance measurement process is done fast enough that the drift from "zero-phase" is negligible. This system allows nearly "hands-off" operation.

The measurement system is mounted on two racks. For radiation studies the refrigerator rack is placed in the Van de Graaff room with the cold head oriented as shown the lower portion of Fig. 1; the control rack is then wheeled into place near the Van de Graaff control console. The 1.75 MeV electron beam passes through a 5 mil aluminum window at the end of the beam tube and then through three foil windows on the cold head before reaching the sample. Preliminary measurements show that for a 10 A beam current we get a current density of about 0.2 A/cm² on the sample. If we assume no energy loss then the dose rate should be approximately 0.1 Mrad/sec. This dose rate is sufficient to cause significant sample heating, especially at low temperatures where the specific heat of the sample is reduced. Currently, we irradiate the crystal with 0.5 sec shots so as to minimize sample heating. At 80 K this should hold the temperature rise to about 1.5 K. Gamma irradiations are carried out by removing the crystal from the cryostat and placing it in a 20,000 Rad/hr ⁶⁰Co cell.

RESULTS AND DISCUSSION

Figure 2 shows the acoustic loss versus temperature spectrum for a Na-swept 5MHz 3rd overtone crystal fabricated from our Toyo bar SQ-B in the as-swept condition and after ⁶⁰Co gamma irradiations of 120 krad and 2.08 Mrad. The data shown in Fig. 2 were taken using the log-decrement method. This particular crystal was plagued by interfering modes; consequently, we are not certain of the production of the 100 K and 135 K Al-hole center peaks. The 23 K peak is present after the irradiations. Figure 3 shows the normalized radiation induced frequency shift versus temperature curves for this Na-swept crystal.

As shown in Fig. 2 the Al-Na center is almost completely removed by the 2.08 Mrad total dose; the other loss peaks such as the 23 K peak all seem to have magnitudes less than about 3×10^{-6} . For temperatures just above the large 53 K Al-Na loss peak the positive frequency shift approaches the expected 200 ppm value. However, the radiation has altered the background temperature dependence so that the shift at room temperature is only about 30 ppm. The shift in the temperature dependence is not understood at this time. The frequency shifts are a "worst case" since this is a Na-swept crystal and most synthetic quartz contains very little sodium. Figure 4 shows that the height of the Al-Na loss peak in this with 10-15 ppm Al-Na as a function of gamma radiation dose. Possibly, the curve could be described by a double exponential.

A 5MHz 5th overtone AT-cut blank from a bar of Sawyer Premium Q quartz from autoclave run D14-45 was Na-swept using our usual procedures. This material has about 0.75 ppm aluminum. The acoustic loss versus temperature spectra for this low aluminum crystal in the as-Na-swept and after several electron irradiations at room temperature are shown in Fig. 5. The peak height of 11×10^{-6} is about twice the value reported by Doherty *et al.*²² for D14-45 quartz Na-swept by a different process. In the current case no Al-OH IR absorption was observed. When comparing the data shown in Fig. 5 with that in Fig. 2 for the 10-15 ppm Al content crystal it is encouraging to note that the 53 K Al-Na center loss peaks scale correctly. Small 135 K Al-Na peaks are also present in this crystal and in the Na-swept Toyo crystal. Toulouse and Nowick²³, who extensively studied these two peaks using dielectric loss measurements on Z-plates, found that the ratio of the low temperature peak, α , to the high temperature peak, β , ranged from 9.7 to 16.8 in Na-swept samples and that the ratio varied depending upon irradiation and thermal treatment. They concluded that the α -relaxation was due to double well jumps on the long-bond side of the Al while the β -relaxation was on the other side. The ratios of the 53 K α -peak to the 135 K β -peak which we see in the acoustic loss spectra are perhaps as large as 100. This large difference is probably caused by the fact that we are looking at the AT-cut thickness shear mode. The model presented by Toulouse and Nowick for the two peaks is different than the earlier one described in Fraser¹⁹ where the β -peak is ascribed to jumps across the Al site from next-nearest-neighbor positions.

Here the irradiations were carried out using 1.75 MeV electrons from the Van de Graaff accelerator. The irradiation geometry was described above and was shown in Fig. 1. The irradiation destroyed the 53 K Al-Na center peak and produced a small peak at about 100 K which is probably related to the Al-hole center. This peak is much smaller than the 100 K peak that we have previously observed in crystals that contain 10-15 ppm aluminum. The irradiation has also produced the 340 K Na-related peak. The size of this peak roughly scales with the Al-content of the crystal when compared to our other samples. Jones and Brown²⁰ have reported the production of a peak near 325 K by x-rays in very dirty fast-growth quartz. One might speculate that the defect responsible for this peak is still the Al-Na center but here the Na has moved one lattice unit away from the central double well location. In a sense this is assigning the old model for the β -relaxation to this newer loss peak. Alternatively, a substitutional germanium may serve as the trap; some of the Ge-related centers show radiation behavior similar to that of the 340 K and 305 K loss peaks. Only a very small 23 K loss peak was observed in this low aluminum sample. A positive radiation induced frequency shift of about 10 ppm was observed for

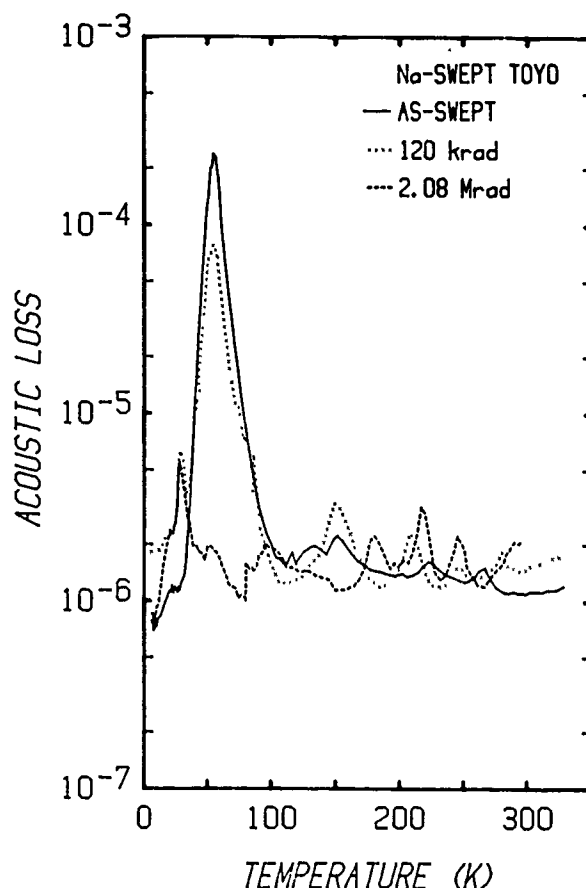


Fig. 2. The acoustic loss of a Toyo SQ-B series crystal is shown as a function of temperature in the Na-swept condition and after two gamma irradiations.

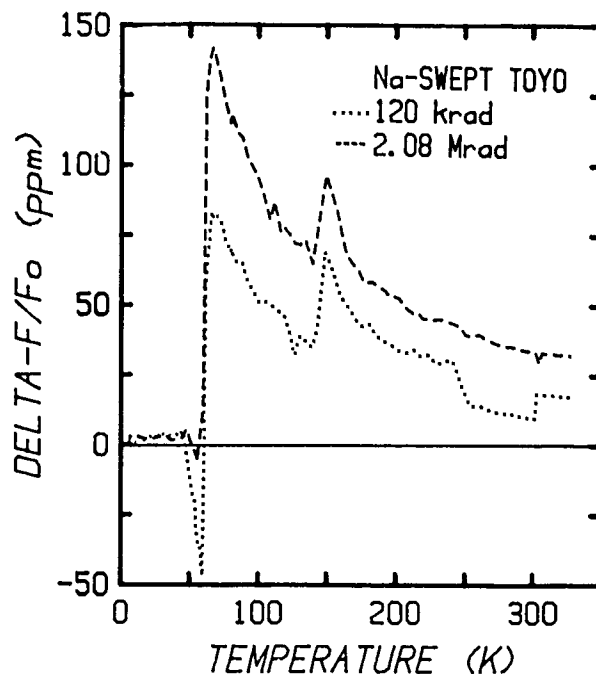


Fig. 3. The irradiated frequency shift, $(f-f_0)/f_0$, where f_0 is the frequency prior to the irradiation is shown as a function of temperature for the Na-swept Toyo crystal.

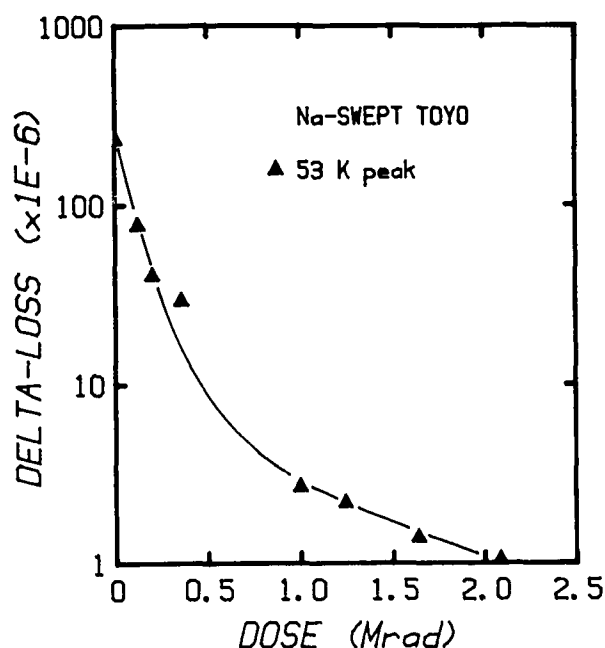


Fig. 4. The height of the 53 K Al-Na loss peak for the Na-swept Toyo crystal is shown as a function of radiation dose. The curve seems to be a double exponential.

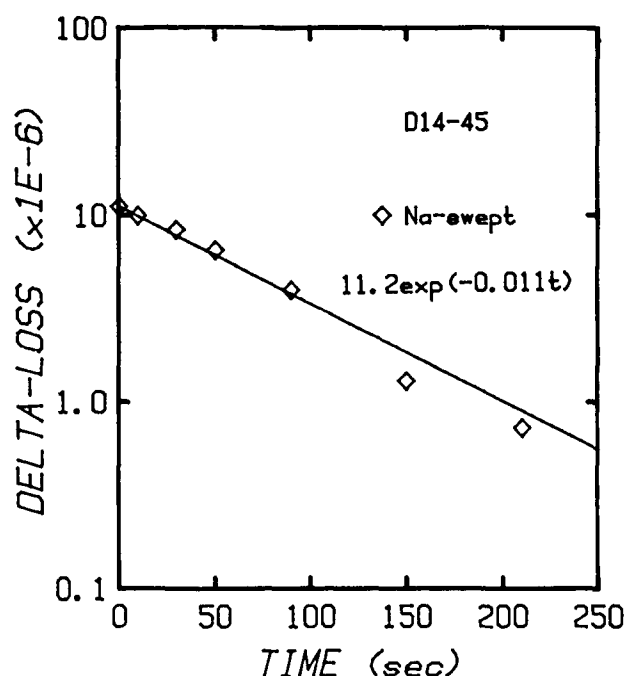


Fig. 6. The decrease in the Al-Na center concentration in the Na-swept D14-45 crystal as measured by the height of the 53 K loss peak is shown as a function of room temperature electron irradiation time. The decrease shows an approximately exponential behavior.

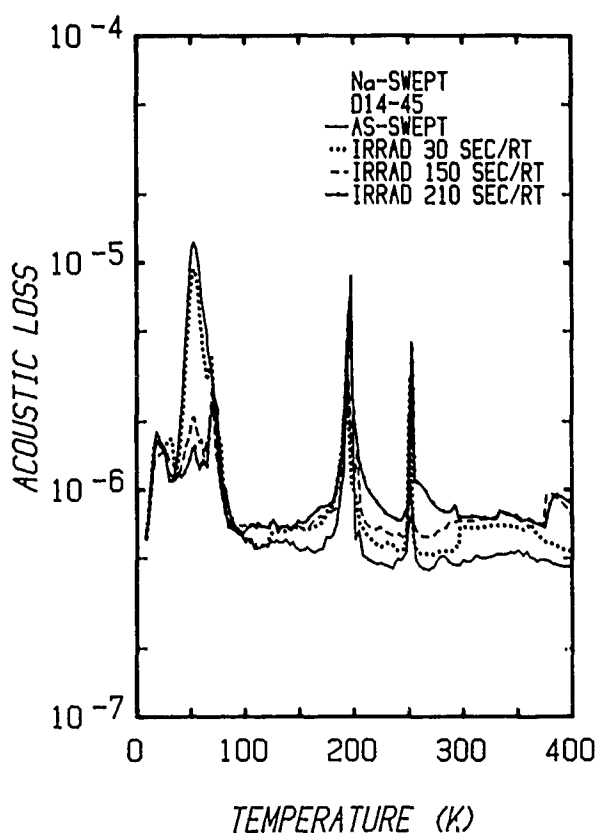


Fig. 5. The acoustic loss versus temperature spectra are shown for a D14-45 series Premium Q quartz crystal in the as-Na-swept condition and after several electron irradiations at room temperature. This material has about 0.75 ppm aluminum. As-grown D14-45 crystals do not show a 53 K Al-Na loss peak.

temperatures just above the loss peak; the magnitude of the shift nearly matches the height of the loss peak. The 1 ppm shift at the turnover temperature was ppm caused by the same change in the background temperature dependence observed for the other Na-swept crystal. Figure 6 shows the reduction of the Al-Na center as measured by the 53 K loss peak versus radiation time for the Na-swept low aluminum sample. The irradiations were carried out at room temperature. We have not yet converted the radiation time into a dose value. The data shown in Fig. 6 are approximately described by $\Delta Q^{-1} = 11 \times 10^{-6} \exp(-0.011t)$ where t is in seconds. Infrared measurements of the Al-OH center versus radiation time usually show a growth described by $(1 - \exp(-Kt))$. This exponential production behavior might be expected since the radiation produces a flood of free electrons and holes which then interact with the Al-alkali and other centers. Hughes²⁴ suggests that about 6×10^{12} electrons/cm³/rad are produced; therefore, we should have 6×10^{17} electrons/cm³/sec produced during our electron irradiations. Thus, at least in the low aluminum samples the production rate may be controlled by the aluminum content.

In our samples with 10-15 ppm the large 53 K Al-Na loss peak tends to mask the radiation produced 100 K loss peak. Figure 7 shows the acoustic loss spectrum of a Li-swept BT-cut crystal. The 5 MHz 3rd overtone blank for this crystal was taken from our Toyo Supreme Q bar SQ-B and Li-swept "in-house". The crystal is estimated to have an aluminum content of 10-15 ppm. No significant 53 K Al-Na peak was observed as shown by the solid curve. A series of electron irradiations were carried out at room temperature. The dotted and dashed curves, for 15 second and 90 second irradiations respectively, show that both the 23 K and 100 K peaks were produced. A negative frequency shift of about 10 ppm was observed at room temperature for this crystal in the fully irradiated condition; this result is slightly larger than one would predict based upon the size of the two loss peaks produced by the radiation. Here the 100 K peak

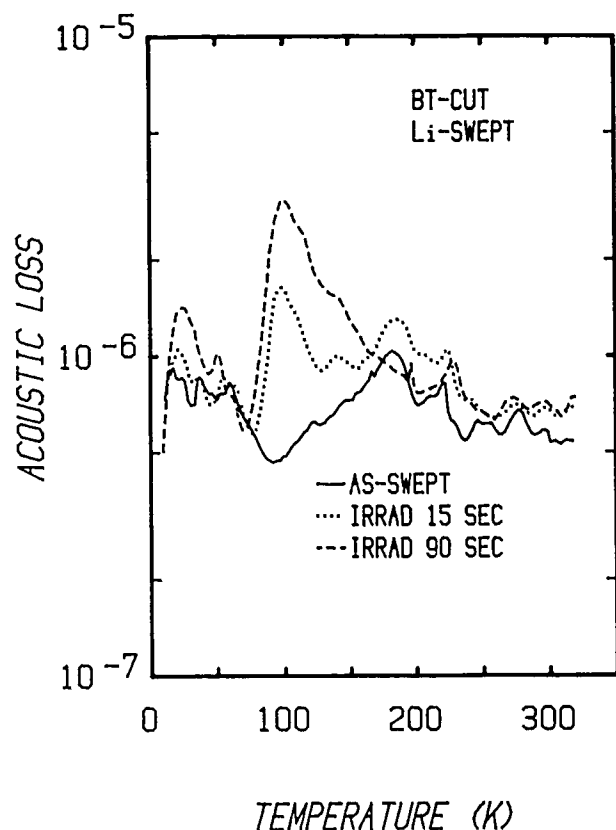


Fig. 7. The acoustic loss spectra of the BT-cut SQ-B series crystal are shown the as-Li-swept condition and after two room temperature electron irradiations.

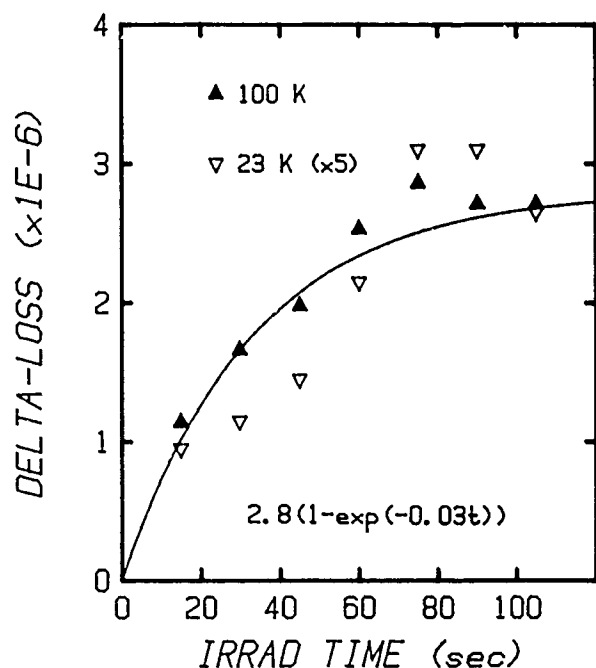


Fig. 8. The growth of the 100 K Al-hole center peak and the 23 K peak are shown as a function of radiation time for the Li-swept BT-cut SQ-B series crystal.

is larger than the 23 K peak; we observe the opposite in similar AT-cut crystals. King²⁵ has pointed out that the 100 K peak is larger in BT-cut crystals than in similar AT-cut crystals. Figure 8 shows the growth of the peak heights as a function of radiation time. The solid curve which was calculated for the 100 K peak is given by $\Delta Q^{-1} = 28 \times 10^{-6}(1 - \exp(-0.03t))$. The much smaller 23 K peak seems to track with the 100 K peak. King¹ has attributed the 100 K peak to the Al-hole center.

The irradiation at room temperature has converted the Al-Li centers into a mixture of Al-OH and Al-hole centers. A subsequent irradiation at liquid nitrogen temperature should convert the Al-OH into Al-hole centers enhancing the 100 K peak. The heavy dotted curve in Fig. 9 shows that an 80 second electron irradiation carried out at 80 K enhances both the 100 K and 23 K peaks. The irradiation was performed in a series of 0.5 second shots so as to minimize sample heating. Both peaks increased in nearly the same ratio.

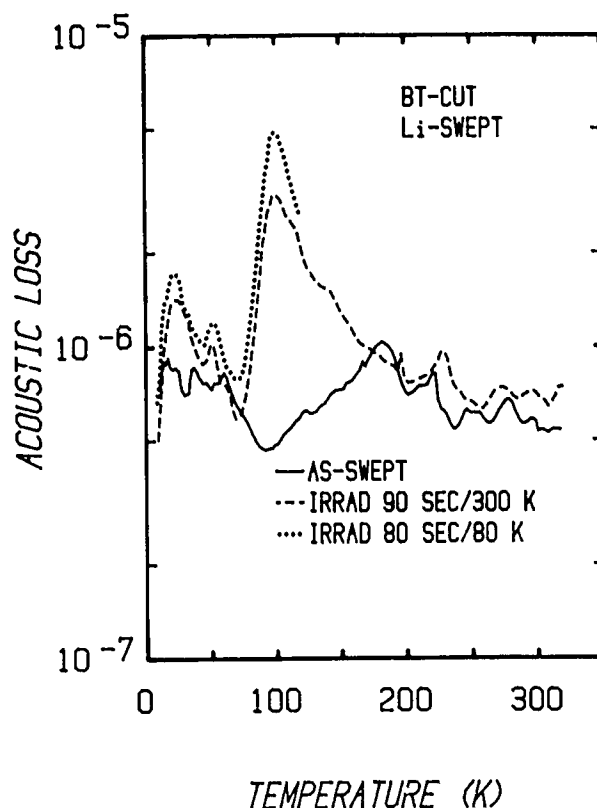


Fig. 9. Both the 100 K Al-hole center peak and the 23 K peak are enhanced by a 80 K electron irradiation.

In H-swept material the aluminum is compensated only by hydrogen forming the Al-OH center. Therefore; it is possible to produce significant numbers of Al-hole centers with an initial irradiation at liquid nitrogen temperature. Figure 10 shows the results for a Toyo SQ-B series AT-cut crystal in the as-H-swept condition, after irradiation at 80 K and after an anneal at 180 K. The solid curve shows the as-swept spectrum. The sample was then cooled to 80 K, electron irradiated, cooled to 8 K, and the dotted curve taken as the crystal was heated to 180 K. Both the 23 K and 100 K peaks were produced by the low temperature irradiation. Here the 100 K peak is slightly larger than the 23 K peak; our previous results for irradiations at room temperature usually show that the 100 K peak is about 1/3 the size of the 23 K peak in AT-cut

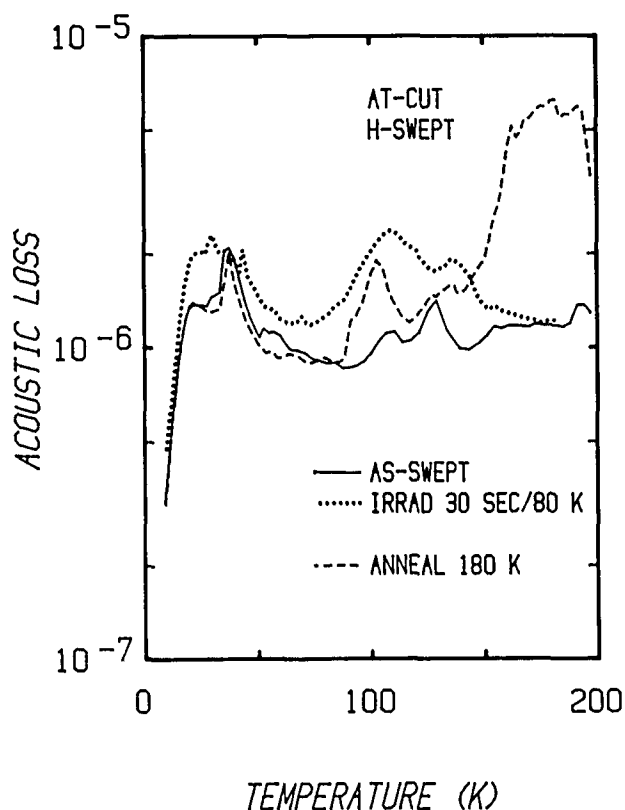


Fig. 10. An initial irradiation at 80 K produced both the 100 K Al-hole center peak and the 23 K peak in this H-swept AT-cut SQ-B series crystal.

samples. After reaching 180 K the crystal was again cooled to 8 K and the dashed curve obtained. The annealing at 180 K (or below) fully removed the 23 K peak and removed most of the 100 K peak. The large upturn in the dashed curve above 150 K is caused by an interfering mode.

The low temperature irradiation results shown in Figures 9 and 10 are consistent with the assignment of the 100 K loss peak to the Al-hole center. Martin¹⁴ showed that both the 23 K and the 100 K peaks when produced by irradiation at room temperature in alkali-swept material anneal out in the same temperature range as the Al-hole center. The results shown above and the annealing data make the assignment of the 23 K to the Al-hole center attractive. However, the production study reported by Martin, Hwang, and Bahadur²¹ showed that in one case the Al-hole center grows more quickly than the defect responsible for the 23 K loss peak. In addition, the 23 K to 100 K peak height ratios for AT-cut samples should not depend upon the temperature during radiation. At the present the 23 K peak should remain unassigned to a particular defect.

CONCLUSIONS

In Na-swept low defect content quartz the Al-Na center follows a single exponential decay with radiation dose. In quartz with a higher aluminum content the center seems to decay with a double exponential dose dependence. A positive frequency shift upon irradiation which tracked with the aluminum content was observed for the Na-swept samples. This shift was caused by the removal of the 53 K Al-Na acoustic loss peak. Because of a change in the temperature dependence of the background frequency the shift at the turn-over temperature was smaller than expected. Irradiation of a Li-swept BT-cut crystal produced both

the 23 K and 100 K loss peaks. The 100 K Al-hole center peak was larger than the 23 K peak; this result is the reverse of our observations in AT-cut crystals. The introduction of these two peaks into the loss spectrum caused the expected negative frequency shift at room temperature. A subsequent irradiation at 80 K of the Li-swept BT-cut sample enhanced both the 23 K and 100 K loss peaks. An initial irradiation at 80 K of a H-swept AT-cut crystal produced both the 23 K and 100 K peaks. Here the peaks had a nearly equal strength; our past observations of room temperature irradiated AT-cut samples showed that the 23 K peak was 3 to 4 times larger than the 100 K peak. The production of the 100 K peak by the initial low temperature irradiation of the H-swept crystal and its enhancement in the Li-swept crystal are consistent with the assignment of the peak to the Al-hole center.

ACKNOWLEDGEMENTS

This work was supported by Sandia National Laboratories. The authors thank D. R. Koehler, J. C. King and T. J. Young for several helpful discussions. J. C. King suggested the measurements on the BT-cut crystals.

REFERENCES

1. Leslie J. Palkuti and Quang T. Truong, Proc. 38th Annual Symposium on Frequency Control, p55.(1984).
2. J. R. Norton, J. M. Cloeren, and J. J. Suter, Proc. 38th Annual Symposium on Frequency Control, p63. (1984).
3. J. Suter and R. Maurer, Proc. 40th Annual Symposium on Frequency Control, in press.
4. T. Flanagan, R. Leadon, and D. Shannon, Proc. 40th Annual Symposium on Frequency Control, in press.
5. J. C. King and H. H. Sander, IEEE Trans Nucl. Sci. **NS-19**, 23 (1972).
6. P. Pelligrini, F. Euler, A. Kahan, T. M. Flanagan, and T.F. Wrobel, IEEE Trans Nucl. Sci. **NS-25**, 1267 (1978).
7. T. J. Young, D. R. Koehler, and R. A. Adams, Proc. of the 32nd Annual Symposium on Frequency Control, p. 34. (1978).
8. D. R. Koehler and J. J. Martin, J. Appl. Phys. **57**, 5205 (1985).
9. J. C. King, Bell System Technical J., **38**, 573 (1959).
10. R. A. Poll and S. L. Ridgway, IEEE Trans Nucl. Sci. **NS-13**, 130 (Dec. 1966).
11. T. M. Flanagan and T. F. Wrobel, IEEE Trans. Nucl. Sci. **NS-17**, 130 (Dec. 1969).
12. B. R. Capone, A. Kahan, R. N. Brown, and J. R. Buckmelter, IEEE Trans. Nucl. Sci. **NS-17**, 217 (Dec. 1970).
13. T. M. Flanagan, IEEE Trans. Nucl. Sci. **NS-21**, 390 (Dec. 1974).
14. J. J. Martin, J. Appl. Phys. **56**, 2536 (1984).
15. W. A. Sibley, J. J. Martin, M. C. Wintersgill, and J. D. Brown, J. Appl. Phys. **50**, 8172 (1979).

16. L. E. Halliburton, N. Koumvakalis, M. E. Markes, and J. J. Martin, J. Appl. Phys. 52, 3565 (1981).
17. L. E. Halliburton, Y. C. Chen, and S. D. Tapp, Proc. 39th Annual Symposium on Frequency Control, p. 259 (1985).
18. L. E. Halliburton, M. G. Jani and R. B. Bossoli, Nucl. Instrum. Methods, 81, 192 (1984).
19. D. B. Fraser, **PHYSICAL ACOUSTICS**, edited by W. P. Mason (Academic Press, New York, 1968), Vol. V. Ch. 2.
20. C. K. Jones and C. S. Brown, Proc. Phys. Soc. 82, 375 (1963).
21. J. J. Martin, Ho B. Hwang, and H. Bahadur, Proc. 39th Annual Symposium on Frequency Control, p. 266 (1985).
22. S. P. Doherty, J. J. Martin, A. F. Armington, and R. N. Brown, J. Appl. Phys. 51, 4164 (1980).
23. J. Toulouse and A. S. Nowick, J. Phys. Chem. Solids, 46, 1285 (1985).
24. R. C. Hughes, Radiat. Eff. 26, 225 (1975).
25. J. C. King, private communication (1986).

CHARACTERIZATION OF BRAZILIAN LASCAS FROM VARIOUS REGIONS
AND THEIR USE FOR SYNTHETIC QUARTZ GROWTH
PART I. LASCAS STUDY

Hideo Iwasaki⁺, Fumiko Iwasaki⁺, Carlos K. Suzuki, Virginia A. R. Oliveira,
Daniele C. A. Hummel, and Armando H. Shinohara

UNICAMP - Instituto de Fisica, 13100 - Campinas, SP, BRASIL

Summary

Brazilian quartz mine survey and lascas sample in situ collection have been carried out in various regions. The lascas quality "control" used is the same adopted since 1940, which is based on a visual classification according to their transparency as "1st", "mix", "2nd", "3rd", "4th", ... lascas, that we call "nominal grades".

In the present research, we report a characterization study of lascas by various techniques, such as X-ray diffraction topography and goniometry, density measurement, ultrasonic attenuation, optical microscopy and atomic absorption spectroscopy.

We have observed that several lascas properties, such as, crystallinity, density, micro cavity concentration and size, crack patterns and ultrasonic attenuation depend on the nominal grades. On the other hand, impurities content (Al, Fe, Ti, Mn, Cu, Cr, Mg, Li) is independent of nominal grades, but it do depend on the extraction sites.

1. Introduction

Since the beginning of this century, Brazil has been the main supplier of natural quartz to international market, as piezoelectric crystals, which reached a peak production during the second world war.

Recently, the interest on lascas has been increased sharply because of their use in synthetic quartz and silica glass production, and large scale consumption to the electronics and telecommunications industries. Only for internal production of metallurgical silicon and silicon-iron alloy, a total of approximately 150,000 tons/year of low graded lascas is used with a prevision of threefold increase in the next few years. The actual exportation rate of brazilian lascas in natura for high tech industries is about 10,000 tons/year.(1)

The quartz in Brazil are found mainly in Minas Gerais (MG), Bahia (BA), and Goias (GO) states. The extraction sites spread in a quite vast regions denominated belts, such as Minas-Bahia belt, Southern Goias belt, Northern Goias belt, and Coastal belt, as represented in Fig. 1. (2)

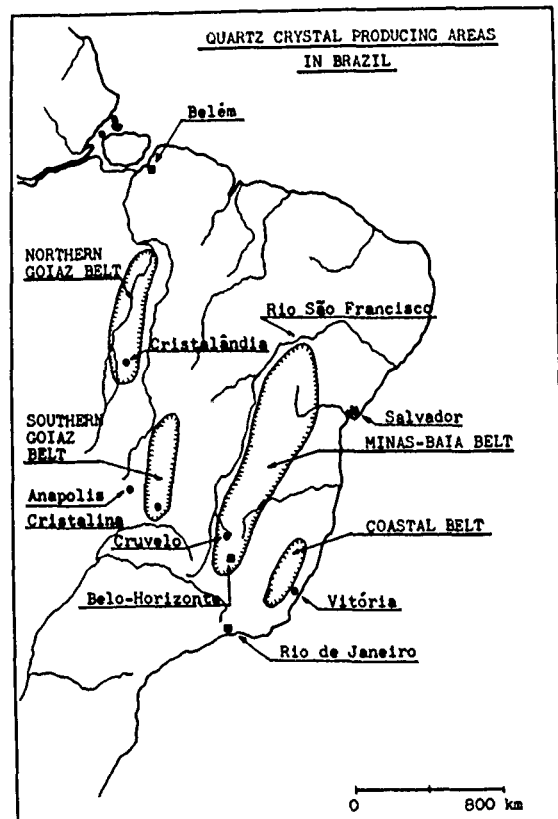


Fig. 1. Principal regions of Quartz production in Brazil. (2)

A typical quartz mine in the region of Diamantina in Minas Gerais state is shown in Fig. 2(a). The classification adopted for long time since 1940, is made only by visual inspections, taking into account of the transparency. In this process, they are classified in several grades, so-called "1st", "mix", "2nd", "3rd", "4th", and so on. Figure 2(b) is a typical view of lasca maker "garimpeiro",

⁺ Japan International Cooperation Agency,
(JICA)



(a)



(b)

Fig. 2 (a) Typical quartz mine in Minas Gerais state. (b) Production of lascas

who peels the quartz block, breaks into small pieces and performs the visual classification. In this situation, lascas graded as "3rd" are believed to be applicable for the synthetic quartz industries in terms of cost and quality performance. But, the main problem is that, until now, lasca classification and quality "control" are relying on "subjective" methods in every extraction sites and "depositos". Therefore, at this moment, the understanding and correlation of grading with the physical-chemical properties of lascas become very important.

In the present research, we have performed a fundamental study of lascas properties characterized by crack patterns, microcavities crystalline perfection and impurity concentration (Al, Ti, Fe, Mn, Cu, Cr, Mg, and Li) and we have studied their correlation with the nominal grades.

2. Characterization Results

Characterization studies by X-ray diffraction topography and goniometry, density measurement, optical micrography, ultrasonics, and impurity analysis by atomic absorption spectroscopy are presented.

2.1. X-ray topography and goniometry

Double crystal X-ray topographic/goniometric system in the non-parallel setting (3,4) with Si (111) asymmetric monochromator and Cu K α radiation was used. This technique permits the use of both (+,-) and (+,+) geometries, as presented in the schematical arrangement of Fig. 3. The utilization of (+,+) geometry is convenient for specimens with high strain field. In the cases of lascas characterization, Bragg case (reflection) topography were used to avoid image blurring caused by strain of

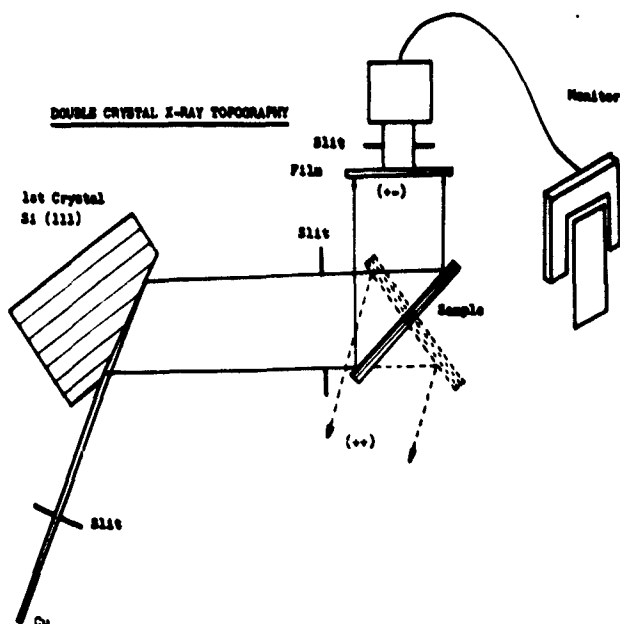


Fig. 3 System of double crystal X-ray topography in Bragg case.

crystal defects observed in the case of transmission topography.

Five pieces for each grade, "1st", "mix", "2nd", "3rd", and "4th" belonging to two mines (three pieces from Diamantina-I and two pieces from Diamantina-II) were collected arbitrarily for topography. All twenty five specimens were treated with a final # 2000 SiC polishing and 10 minutes chemical etching in HF solution.

Figure 4 shows a typical topographic image of lascas graded as "1st", "mix", ..., "4th" from Diamantina-I mine for (0003) reflections on (+,-) geometry. The corresponding rocking-

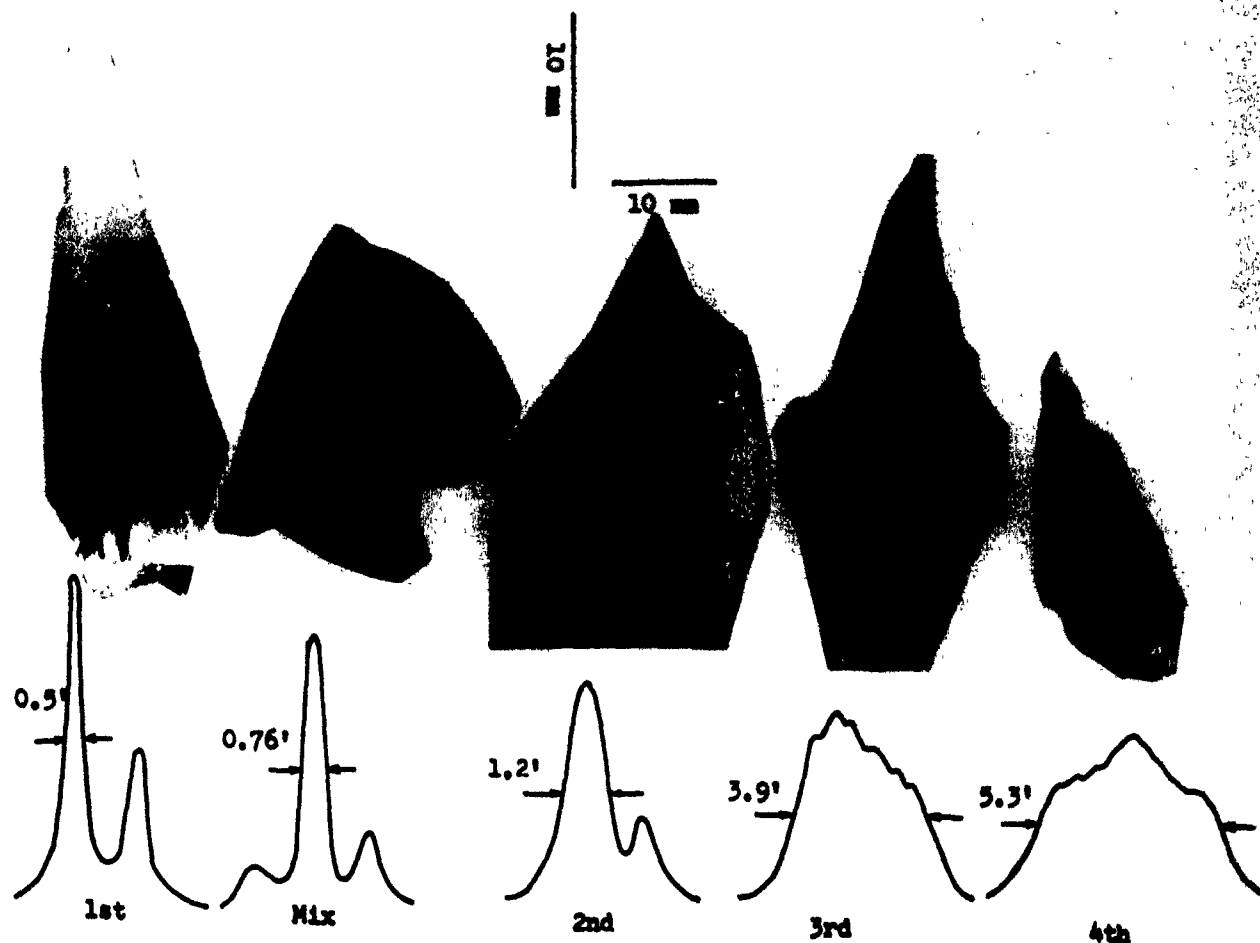


Fig. 4 X-ray topography and diffraction profiles of graded lascas; Z-cut, (0003)-Bragg case.

curves drawn in the same figure furnish the quantitative measurement of average strain caused by the imperfections, such as boundaries, microcavities and crack pattern. All topographs were taken at the peak position of the respective diffraction profiles.

It can be observed through the topographs that the "1st" grade lasca is homogeneous with out boundaries. The boundaries occur in "mix" lascas and continuously increase through "2nd" lascas to lower grades. Therefore, the area delimited by the boundaries which we call "grain" becomes smaller and smaller as the grade goes down.

A comparison of topographic image and optical photograph taken on the same specimen reveals a direct correlation between "crack" and "boundary", but the inverse is not always valid. The X-ray images show an interesting property about the propagation directions of boundaries, which preferentially run along the crystallographic a_1 , a_2 , and a_3 directions.

The half width of maximum (H.W.) of the diffraction profiles for the twenty five specimens from Diamantina-I and Diamantina-II mines

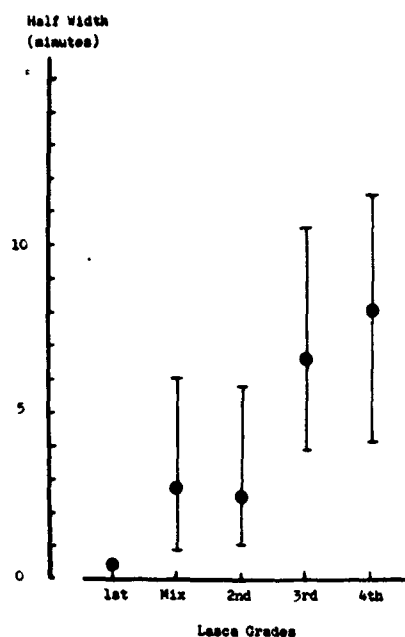


Fig. 5 H.W. of rocking curves of graded lascas

are presented with respect to their grade in Fig. 5. The solid circles represent the average value which shows a grade dependent behaviour. It can be noted that there is a dispersion of these values from "mix" through "4th" lascaras with the exception of "1st" lascaras

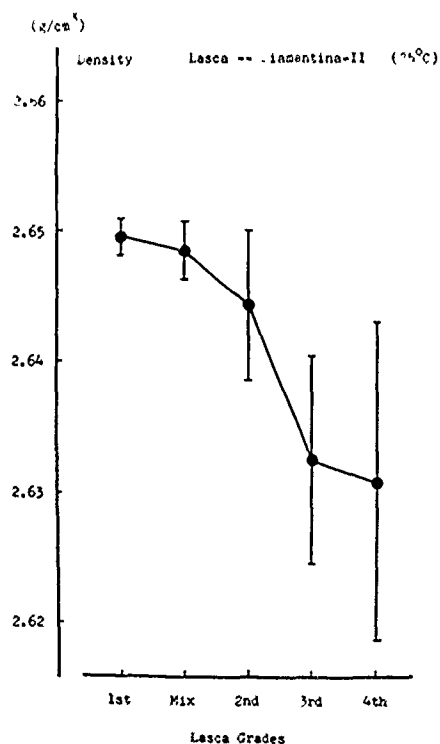
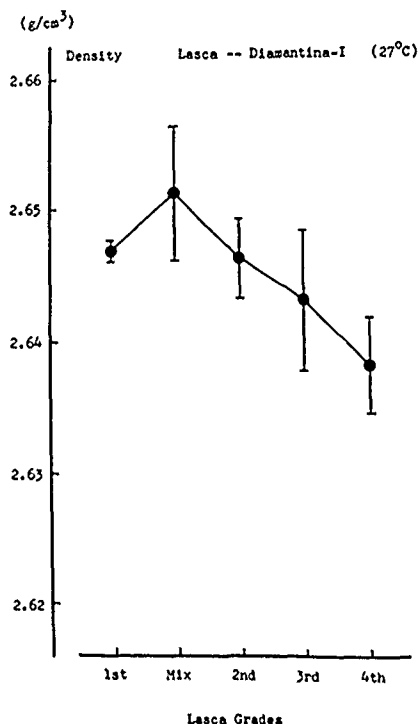


Fig. 6 Density characterization of graded lascaras.

which are very homogeneous. For the H.W. values in the interval 4 - 6 minutes of arc, there is a superposition of the four grades of lascaras ("mix" to "4th"), which means an ambiguity deviation of the classification method in terms of crystalline imperfections.

2.2. Density measurement

Lascaras extracted and "as-classified" from various mines were characterized by density measurement using the Archimedean method (4). Five pieces of lascaras for each grade ("1st" to "4th"), in a total of twenty five specimens, were collected from each quartz mine.

The results of these measurements for Diamantina-I and Diamantina-II lascaras show a typical relationship between the average density and lascar grades as represented in Fig. 6. As it can be observed, a range of deviations in density values is superposed for several grades.

Such characteristics are usually observed in all cases of lascaras, independently of the origin and the types of mines. The density decreasing through lower grades ("1st" to "4th") is in correlation with the crack patterns of lascaras, which become more and more intense, especially for "3rd" and "4th" lascaras. The figure 7(a) shows the crack patterns in lascaras graded as "2nd" and "4th", where the preferred orientations of crack propagation can be noticed.

The density characterization has the advantage of being a very simple and fast measurement and also the samples can be used directly in the as-collected state avoiding any kind of preparation process, such as cutting and polishing. Density data in the practical point of view is probably important for hydrothermal growth to correlate with the solubility rate of lascaras as the nutrient. For quality control of lascaras, it is useful for the first step of standardization method.

2.3. Observation of microcavities

Microcavities observations were performed in lascaras from Diamantina regions. In a similar way as previous X-ray characterization, five pieces for each grade, three from Diamantina-I and two from Diamantina-II, in a total of twenty five samples for two mines were prepared by cutting Z-plates with 1 mm thickness and polishing to get transparency. Figure 7(b) shows the structure of a microcavity, which occurs preferentially along the boundaries.

Counting and classification of microcavities by optical microscopy have been performed according to the interval of their sizes and characterized by "classes", 0-20 μm , 20-40 μm , 40-60 μm , ..., 100-120 μm , and bigger than 120

μm . For size estimation, the longer dimension of the cavity has been adopted. The results listed in Table I show a sharp increase of small microcavities for lower grade lascar, particularly for 0-20 μm class. Histograms of microcavity density in graded lascar are presented in Fig. 8.

The quantitative analysis of the microcavity density reveals an exponential behaviour and fits in an equation of the type

$$N = A \exp (B \cdot X),$$

where X corresponds to the number of "classes" (0,1,...,5) of the size interval and A and B are constants that can be obtained from measured data. Especially for "3rd" and "4th" lascar, the exponential law is very well fitted.

Table I Classification of microcavity densities, (number/cubic cm)

Class	Size (μm)	1st*	Mix	2nd	3rd	4th
0	0 - 20	800	1820	1856	6020	29300
1	20 - 40	40	346	456	1358	5370
2	40 - 60	-	251	167	738	2912
3	60 - 80	-	95	43	258	602
4	80 -100	20	98	33	84	102
5	100 -120	-	70	13	22	30
6	120<	-	33	2	4	-
Total		860	2713	2570	8484	38316

* Microcavities were found out in only one sample within 5 samples prepared.

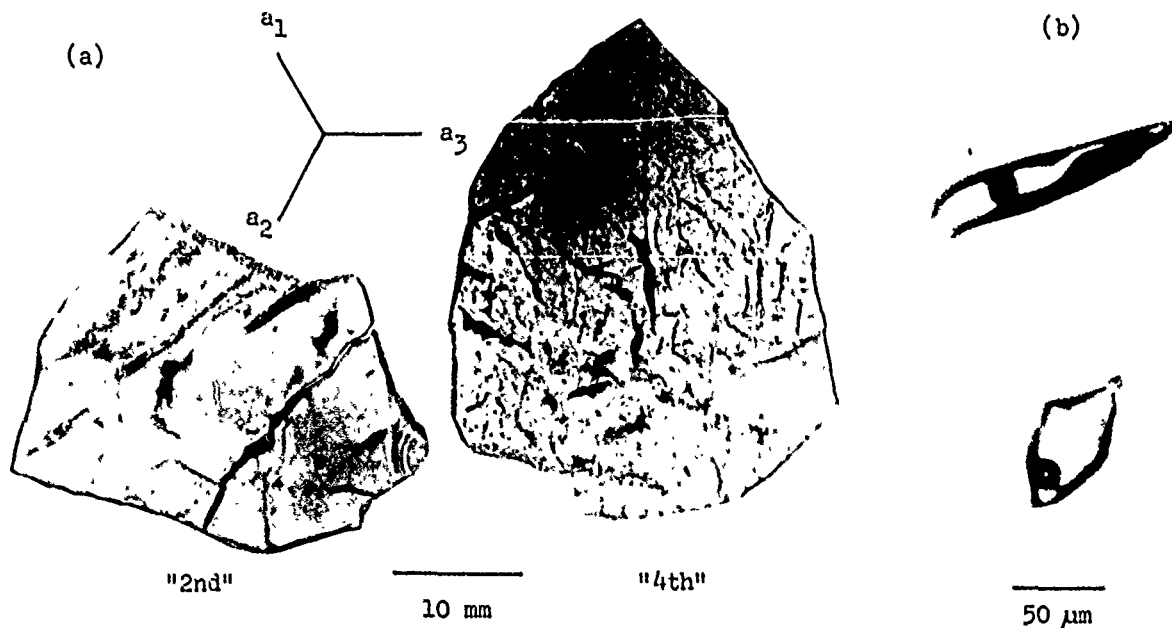


Fig. 7 (a) Crack patterns of lascar graded as "2nd" and "4th". (b) Microcavities

2.4. Ultrasonic attenuation

Ultrasonic characterization of graded lascar was performed by measuring the attenuation of longitudinal wave in the frequency of 10MHz by the pulse echo method (6). The specimens were cut from lascar in a form of plates with a normal approximately in the Z-direction. The results show a correlation between the grades and the attenuation of the longitudinal wave, as shown in Fig.9. The solid circles in the plot of Fig. 9 represent the average values of five samples collected from two mines. This shows clearly that crack patterns, which are grade dependent, influence the ultrasonic attenuation characteristics in lascar.

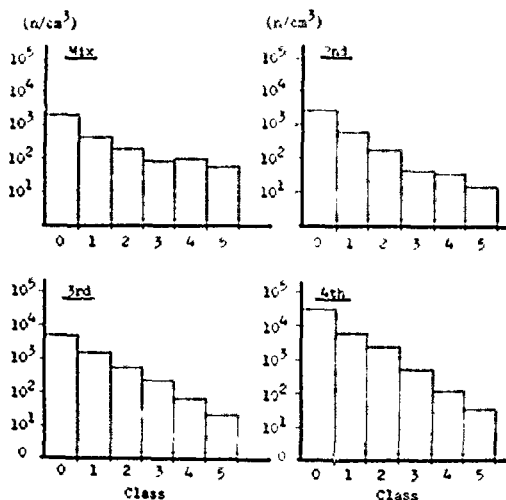


Fig. 8 Histograms of microcavity densities.

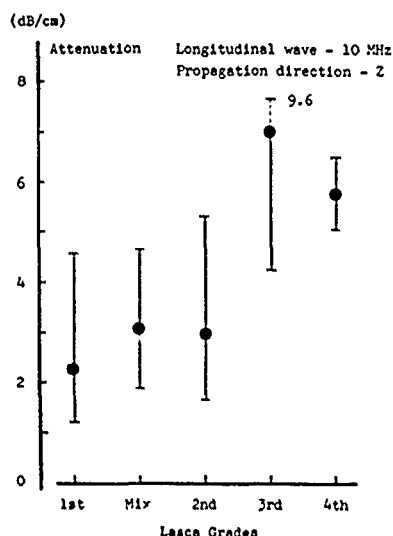


Fig. 9 Ultrasonic attenuation of graded lascas.

2.5. Impurity analysis by Atomic Absorption Spectroscopy (AAS)

The impurity analysis results of synthetic quartz have been published (8-12). For example, Fluesmier has communicated his results on impurity analysis of natural, synthetic quartz and silica glass. But we have not been able to find yet a detailed publication of impurity characterization of graded lascas.

In the present research, impurity characterization of eight elements, Al, Fe, Ti, Cr, Cu, Mn, Mg, and Li, in lascas have been performed by AAS (Model AA-670, Shimadzu). The specimen preparation in form of solution and the standard solutions for calibration followed standard process, taking care to avoid contamination. In order to keep a high reliability of the analysis with respect to the fluctuations due to instrument, preparations of solutions and calibration, we have developed and used the "self consistent process" in analytical operation. (13)

Typical results of the analysis in two sets of graded lascas of "1st", "mix", "2nd", "3rd", and "4th" from Diamantina-I and Diamantina-II mine are presented in Table II. We can note that Al, Fe, Mg, and Li, are nearly independent of grades. Other elements, Ti, Cu, Cr, and Mn have not been detected. In the case of Diamantina-I lascas, the "mix" lascas present a higher impurity concentrations than other grades.

Impurity concentration of "3rd" grade lascas extracted from various regions are listed in Table III. Among the impurities analyzed, we can observe that Al and Fe are the dominant impurities. We emphasize that they present a remarkable variation from mine to mine. A "Geological Map" of impurity contents, Al and Fe,

is prepared preliminarily on lascas extracted from mine in "Minas-Bahia Belt", as show in Table IV.

Table II Impurity contents in lascas exploited from mines of Diamantina-I and -II.

Impurity (wt ppm)	Grade	L a s c a s	
		Diamantina-I	Diamantina-II
Al	1st	23.9	23.1
	Mix	58.5	28.4
	2nd	48.0	31.2
	3rd	30.1	24.5
	4th	28.5	22.7
Fe	1st	5.6	9.9
	Mix	10.5	11.9
	2nd	4.7	11.3
	3rd	4.4	11.9
	4th	4.3	11.5
Mg	1st	0.4	0.9
	Mix	3.3	0.8
	2nd	1.1	1.0
	3rd	0.7	0.8
	4th	0.4	0.7
Li	1st	0.4	0.7
	Mix	1.3	0.8
	2nd	1.3	0.9
	3rd	0.5	0.8
	4th	0.7	0.6

Ti, Cu, Cr, and Mn : not detected

Table III Impurity contents comparison of "3rd" graded lascas.

Lascas (3rd Grade)	Al (wt ppm)	Fe (wt ppm)	Mg (wt ppm)	Li (wt ppm)
Diamantina-I	30.1	4.4	0.7	0.5
Diamantina-II	24.5	11.9	0.8	0.8
Ribibiu	47.2	31.5	0.9	1.1
Ouro Fino	195.	4.8	0.2	2.0
Bicas	241.	8.2	0.2	26.5
Brumado	25.9	7.4	1.3	0.2
Cabral	117.	10.8	0.3	4.9
Oliveira dos Brejinhos-I	15.3	4.8	0.7	0.8
Oliveira dos Brejinhos-II	64.2	3.6	0.2	2.0

Tabel IV Preliminarily prepared "Geological Map" of impurity contents in lascas exploited from mines in "Minas - Bahia Belt".

Lascas	Al					Fe				
	1st	Mix	2nd	3rd	4th	1st	Mix	2nd	3rd	4th
Diamantina-I	23.9	58.5	48.0	30.1	28.5	5.6	10.5	4.7	4.4	4.3
Diamantina-II	23.1	28.4	31.2	24.5	22.7	9.9	11.9	11.3	11.9	11.5
Ribibiu				47.2					31.5	
Bicas				214.					8.2	
Brumado				25.3					7.4	
Cabral				117.					10.8	
Cruvelo				51.3					15.1	
Ouro Fino				195.					4.8	
Oliveira dos Brejinhos-I				15.3					4.8	
Oliveira dos Brejinhos-II				64.2					3.6	

3. Discussion and Conclusion

With the objective of studying the physical-chemical fundaments of Brazilian graded lascas, characterization studies have been conducted by X-ray topography and goniometry, density measurement, optical micrography, ultrasonic analysis and impurity analysis by AAS.

Principal results obtained are;

- The crystallinity, structural defects, crack configurations and microcavity concentration depend on the nominal grades ("1st", ... "4th").
- The impurities concentration, on the other hand, are independent of the grades for eight impurity elements analyzed, but depend on the individual mine.

In the process of lascas classification from "1st" to "4th", they are usually taken from a quartz block. This may justify there is no significant variation in the impurity content among the different grades. However, the visual transparency depend basically on the cracks configuration (except the milky quartz) which is the main parameter for the grading process used by "garimpeiros".

Therefore, the following questions remains; "Why parts of a same crystalline quartz block found in nature have different crack configurations, even though they are equivalent in impurity quality?"

This can be answered through the analysis of crystalline perfection of a natural faceted block of quartz, as we can notice, almost always the bottom is opaque, and it becomes gradually more transparent to the top direction. This means that the crystal perfection is improved through the growing on the Z-direction. A previous study by X-ray topography and goniometry reveals in detail this improvement process occurring in natural growth. (14) One probable explanation of the phenomenon would be the growth model containing at the beginning many microcrystals orientated at random and the formation of grain boundaries as they grow. As the crystal continues growing, a selection of preferential orientation of small crystallites, which act as seeds, takes place. Consequently, there is a crystal quality improvement through the growing direction on Z-axis. (15) The cracks could be induced by the strain caused by crystallites misorientation. Crack configuration and concentration are regarded as a function of the intensity of strain field. This model explains that the small dimensions of the grains and the high concentration of crack in "3rd" and "4th" lascas are associated with a region near the bottom with a great number of micro-seeds.

The establishment of fundamental criterion for classification and quality control of Brazilian lascas (13) becomes possible with the present study. The knowledge of basic properties of lascas is also very important for cost and quality improvement of synthetic quartz and silica glass productions. In addition, these results suggest the possibility to establish the standardization of lascas control.

Acknowledgment

The authors would like to thank Dr. J. Yoshimura, Yamanashi University, for his help to construct the X-ray topographic system, and Professors K. Kohra and S. Kikuta, University of Tokyo, for supplying the high precision goniometers. Professor D. G. Pinatti, UNICAMP, permitted us to use the ultrasonic equipments.

Financial supports by FAPESP and CNPq are also acknowledged.

References

- (1) "Sumário Mineral", Departamento Nacional da Produção Mineral, 5, 86 (1985)
- (2) R.E. Stoiber, C. Tolman and R.D. Butler, "Geology of Quartz Crystal Deposits", American Min. 30, (1945) 245
- (3) K. Kohra, M. Yoshimatsu and I. Shimizu, "Direct Observation of Imperfection in Crystals", Ed. by J.B. Newkirk and J.H. Wernick, Interscience, (1962) 461
- (4) J. Yoshimura, Y. Koishi and C.K. Suzuki, "Growth Defect on Natural Beryl Crystals" J. Crystal Growth, 73, (1985) 275
- (5) D.C.A. Hummel, H. Iwasaki, F. Iwasaki and C.K. Suzuki, "Density determination of Quartz Lascas", Presented at 29th Brazilian Ceramic Congress, Criciúma, SC, April 1985, To be published at Revista Cerâmica
- (6) A.H. Shinohara, F.O. Rangel, H. Iwasaki, D.C.A. Hummel and C.K. Suzuki, "Ultrasonic Characterization of Natural and Synthetic Quartz Crystals", Presented at the 29th Brazilian Ceramic Congress, Criciúma SC, April 1985, To be published at Revista Cerâmica
- (7) A.H. Shinohara, H. Iwasaki, C.K. Suzuki, F. Iwasaki and D. Torikai, "Ultrasonic Characterization of Solids, Part I Synthetic and Natural Quartz", Presented at the 30th Brazilian Ceramic Congress, Rio de Janeiro, April 1986. To be published at

- (8) A.L. Fluesmier, "Trace Analysis in Quartz" Report 1984, RADC-TR-84-51, Order N^o AD-A142425, 20 pp (Eng.) Avail. NTIS, from Gov. Rep. Announce Index (U.S.) 1984, 84 (20), 64
- (9) J. Yoshimura, T. Miyazaki, T. Wada, K. Kohra, M. Hosaka, T. Ogawa and S. Taki, "Measurement of Local Variations in Spacing and Orientation of Lattice Plane of Synthetic Quartz", J. Crystal Growth, 46, (1976) 691
- (10) A.F. Armington, J.F. Balascio, "The Growth of High Purity, Low dislocation Quartz", Proc. 38th Ann. Freq. Control Sympos, (1984) 3
- (11) B. Sawyer and D.R. Kinlock, "Cultured Quartz of Low Aluminum Content from Production Size Autoclaves", Proc. 38th Ann. Freq. Control Sympos. (1984) 8
- (12) A.F. Armington and J.F. Balascio, "The Growth of High Quality Quartz in Commercial Autoclave", Proc. 39th Ann. Freq. Control Sympos. (1985) 230
- (13) C.K. Suzuki, H. Iwasaki and F. Iwasaki, "Brazilian Quartz for Electronic and Telecommunications Industry", Presented at 2nd Symposium on Electronic and Telecommunications Industry Materials, June 1986, Campinas, SP
- (14) C.R.L. Farias, F.Iwasaki and C.K. Suzuki, "Perfectness of Natural Quartz Block by X-Ray Topography", Presented at 30th Brazilian Ceramic Congress, April 1986, Rio de Janeiro. To be published at Revista Ceramica
- (15) D.P. Grigoriev, Ontogeny of Minerals, Israel Programme for Sci. Transl., Jerusalem (1965)

CHARACTERIZATION OF BRAZILIAN LASCAS FROM VARIOUS REGIONS
AND THEIR USE FOR SYNTHETIC QUARTZ GROWTH

PART. II PROPERTIES CORRELATION

Carlos K. Suzuki, Armando H. Shinohara, and Virginia A.R. Oliveira
UNICAMP - Instituto de Física, 13100 - Campinas, SP, Brazil
and
Simão Takiya and Jozsef Kiss
ABC-Xtal Microeletrônica S.A.
Av. Brasil, 20201 - Rio de Janeiro, Brazil

Summary

Hydrothermal growth runs in production au clave were carried out using nutrient lascas of various mines spread over quartz production belts in Brazil.

Impurity content, in particular Aluminum, in lascas was correlated with the synthetic crystals properties. The effect of lascas grade on the quality of grown crystals was also studied by X-ray diffraction topography, ultrasonic attenuation, inspectoscopy, infrared absorption, and atomic absorption technique(AAS).

1. Introduction

The fundamental aspects of Brazilian lascas properties, the understanding of their classification and grading process, and the dependence of impurities content with the extraction sites were presented in the Part I. (1)

With respect to synthetic quartz growth, low Al content crystals have been developed by improving the growth conditions. (2) (3)

In the present research, lascas extracted from various regions and previously characterized were used to grow crystals in commercial autoclaves of the cultured quartz plant in ABC Xtal, Rio de Janeiro.

The synthetic quartz properties have been correlated with the nutrient lascas characteristics. Fundamental point to be considered is to verify the effect of lascas grade on the synthetic quartz quality, even though "3rd" grade lasca is supposed to be convenient and traditionally being used for large scale quartz crystals production.

Therefore, in the present research, three types of lascas graded as "2nd", "3rd", and

"4th" extracted from the same mine in Diamantina region were used to grow synthetic quartz crystals. In addition, "3rd" graded lascas extracted from various mines in "Minas - Bahia Belt" were used.

2. Crystal Growth

Hydrothermal growth of Y-bar quartz crystals has been conducted according to the growth conditions listed in Table I. Average growth rate on the Z-direction can be estimated as 0.7 mm/day.

Table I Growth conditions

- | |
|---|
| 1) Growth Temperature = 345°C, |
| 2) Temperature Difference = 15 ~ 35°C |
| 3) Pressure = 1700 kg/cm ² |
| 4) Days of Growth = 35 ~ 40, |
| 5) Solution = NaOH (1.0 mol) +
Li ₂ CO ₃ (0.025 mol) |
| 6) Autoclave, Inner Diameter = 25 cm,
Length = 600 cm. |

3. Characterization Results

3.1. Impurity analysis by AAS

The sample solution preparation process and the measuring methodology adopted have been previously discussed. (1) The analytical results of synthetic quartz grown from eight runs are shown in Table II, together with the nutrient lascas characteristics. The values of Al and Fe contents for all regions (whole) and Z-regions are listed. The influence of unlined autoclave on Fe content is evident.

Typical Al contents in lascas ranging 15 to 60 wt ppm and exceptionally high values of

241 wt ppm (Bicas) and 195 wt ppm (Ouro Fino) were also used. However, impurity content of grown crystals on the whole region depend on the extraction sites of lascas. It is observed that Al content in the Z-region does not show a significant difference for the grown crystals, even though the original lascas were quite different, being an order of magnitude higher in some of them.

Table III presents the analysis of three types of synthetic crystals grown using lascas graded as "2nd", "3rd", and "4th" from the same mine, Diamantina-I. In this case, we can observe that the impurities content of the nutrient lascas does not influence significantly on the quality of synthetic crystals for commercial use. This result is quite different from the previous experiments using lascas from different mines.

3.2. X-ray topographic observations

The X-ray topographic characterization was performed by using the double crystal non-parallel setting discussed in part I (1), where the reflection geometry was adopted. For the synthetic crystals characterization, the transmission (Laue case) geometry was also used, as represented in Fig. 1.

Topographic images for a set of four Y-cut samples for (20 $\bar{2}$ 0) and (0003) reflections are shown in Fig. 2. The influence of Al impurity is notorious on the s-region dimension in detriment of Z-region size. This is in accordance with a previous result observed in Al-doped synthetic quartz. (4) The (0003) reflection topography reveals the contrast of growth striations, supposed to be fluctuations of impurity concentrations, which become more and more intense with the Al content in lascas. We can also observe from the topographs that the high Al content in lascas induces the generation of dislocations from the seed surface. This is distinct from the case of low Al content, where dislocations extend from the dislocation in the seed.

The reflection (Bragg case) topography as shown in Fig. 3, on the other hand, reveals a qualitative difference on lattice parameters in the growth regions. A set of three topographs was taken in three positions of the rocking curve of (02 $\bar{2}$ 0) reflection as shown in the figure. The contrast inversions of +X and s-regions, for example, in the topographs (2) and (3), means that in the angular position (3), +X-region is reflecting and in position (2) practically it does not reflect. The reciprocal is valid for s-regions.

Table II Comparison of impurity contents, Al and Fe, in "3rd" graded lascas and grown quartz crystals.

Samples	Lascas (3rd grade)		Synthetic Quartz			
	Al	Fe	Al		Fe	
			whole	Z-region	whole	Z-region
Diamantina-I	30.1	4.4	22.3	11.1	5.7	4.9
Bicas	241.	8.2	77.5	10.8	22.2	6.5
Cabral	117.	10.8	26.7	15.9	20.4	7.3
Brumado	25.9	7.4	22.9	10.2	13.8	5.3
Cruvelo	51.3	15.1	14.7	8.3	6.5	6.8
Ouro Fino	195.	4.8	72.0	17.9	4.1	6.0
Oliveira dos Brejinhos-I	15.3	4.8	19.9	7.0	5.5	6.3
Oliveira dos Brejinhos-II	64.2	3.6	25.5	11.6	5.3	7.4

unit:wt ppm

Table III Comparison of impurity contents in synthetic quartz grown from lascas graded as "2nd", "3rd", and "4th" extracted from a mine, Diamantina-I.

	Lascas -- Diamantina-I (wt ppm)					
	2nd grade		3rd grade		4th grade	
	Lasca	S. Q. (whole)	Lasca	S. Q. (whole)	Lasca	S. Q. (whole)
Al	48.0	18.7	30.1	22.3	28.5	16.9
Ti	-	-	-	-	-	-
Fe	4.7	3.0	4.4	5.7	4.3	4.1
Mn	-	-	-	-	-	-
Cr	-	-	-	-	-	-
Cu	-	-	-	-	-	-
Mg	1.1	0.8	0.7	0.9	0.4	1.1
Li	1.3	1.6	0.5	1.5	0.7	2.0
Total	55.1	24.1	35.7	30.4	33.9	24.1

S. Q. = Synthetic quartz, (-) = not detected.

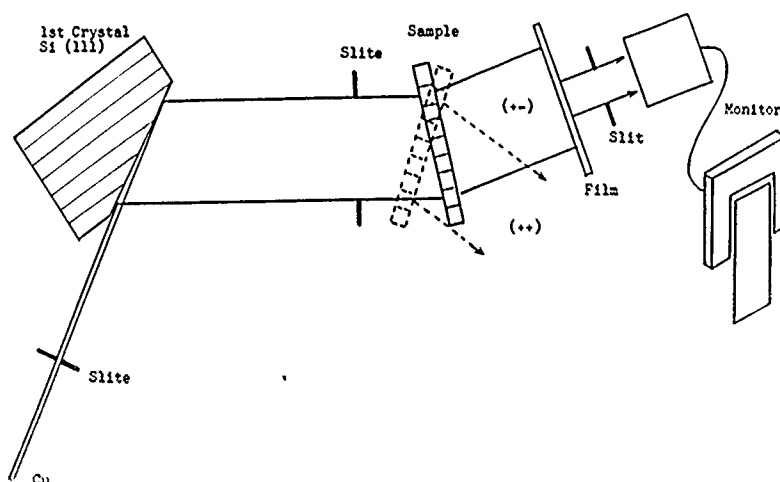


Fig. 1 System of Double Crystal X-ray Topograph-Laue case

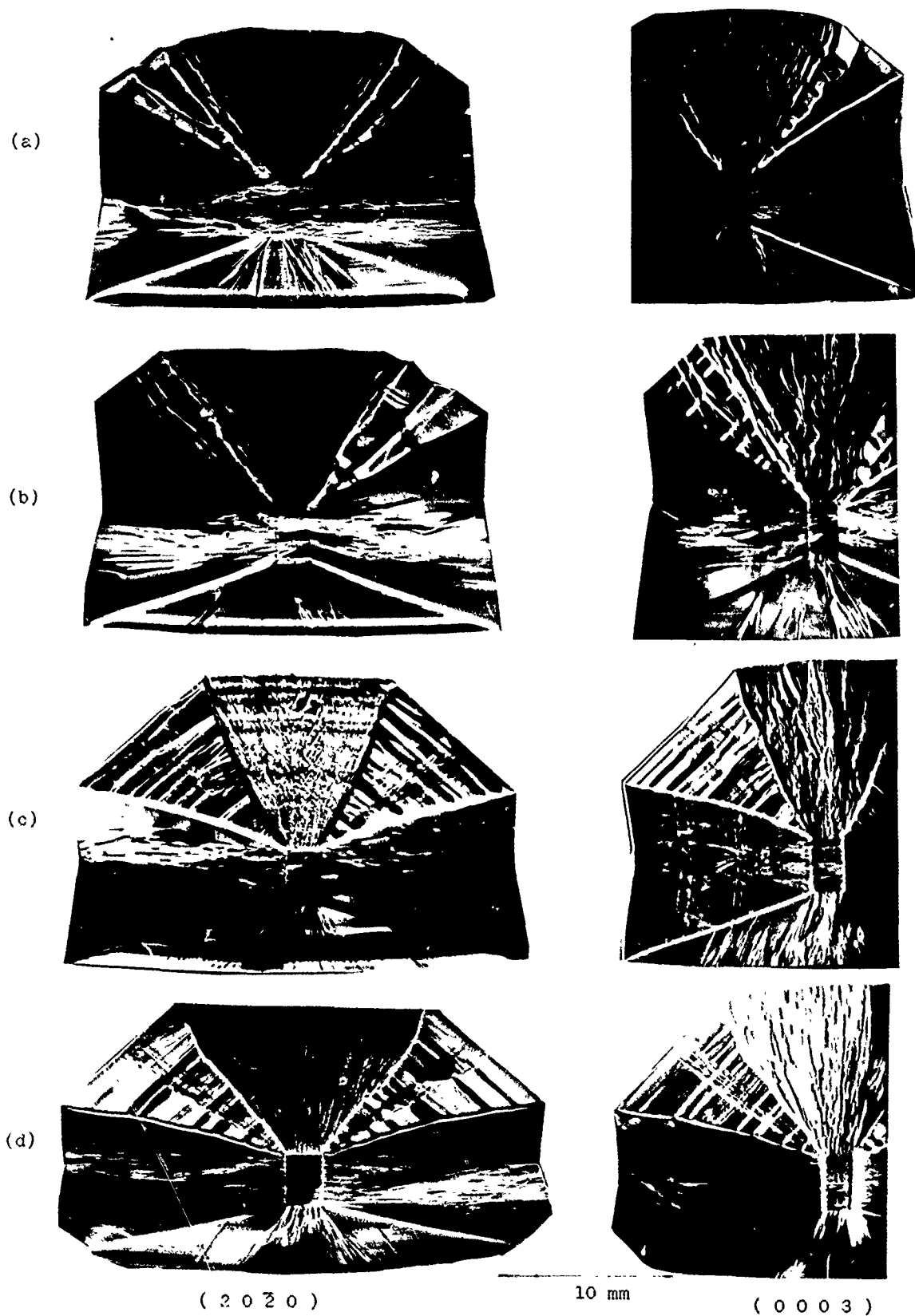


Fig. 2 X-ray topographs of Y-cut plates, Laue case. (reserved print)

- | | | |
|-----|---|----------------|
| (a) | Synthetic quartz grown by lascas from Diamantina-I, | (Al = 48 ppm) |
| (b) | Synthetic quartz grown by lascas from Cabral, | (Al = 117 ppm) |
| (c) | Synthetic quartz grown by lascas from Bicas, | (Al = 241 ppm) |
| (d) | Synthetic quartz grown by lascas from Ouro Fino, | (Al = 195 ppm) |

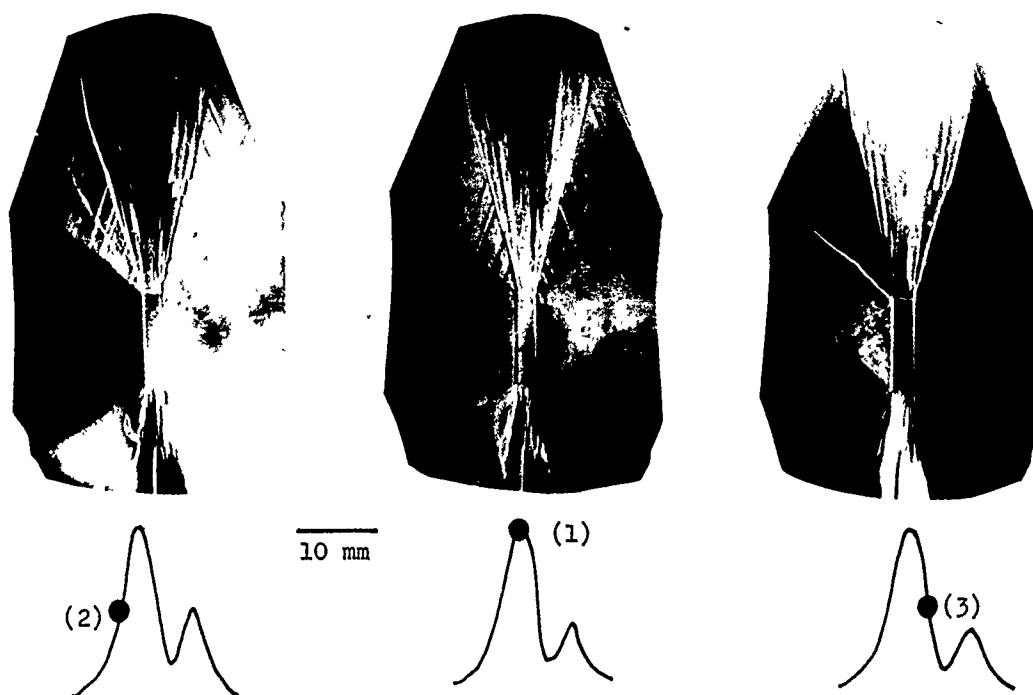


Fig. 3 X-ray topographs of Bragg case. (reversed print)
Synthetic quartz grown by Ouro Fino lascas (Al= 195 ppm)

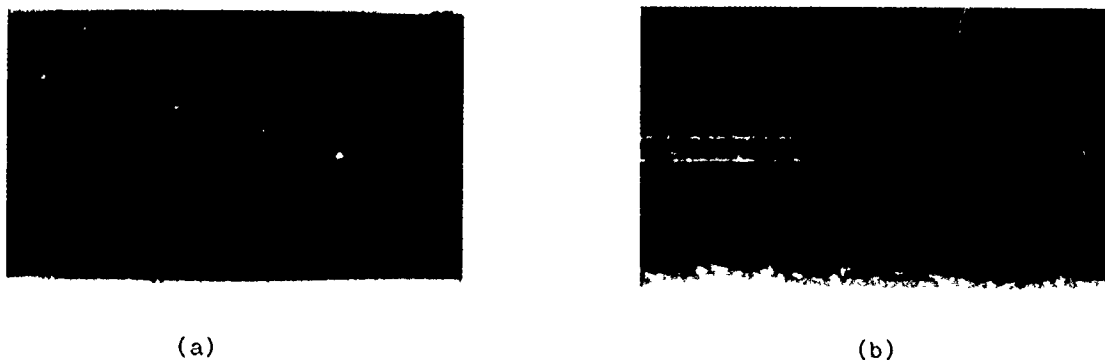


Fig. 4 Distribution of inclusions.

- (a) Synthetic quartz grown by lascas from Diamantina-I, (Al= 30.1 ppm, Fe= 4.4 ppm)
(b) Synthetic quartz grown by lascas from Bicas, (Al=241. ppm, Fe= 8.2 ppm)

Therefore, for the angular positions of (2) and (3), the variation of lattice parameters between s- and +X-regions can be roughly estimated as $\Delta d = 0.001 \text{ \AA}$.

3.3. Optical inspectoscopy

Optical inspectoscopy observations of inclusions in synthetic quartz show an increase of concentration with the Al content in nutrient lascas. As an example, Fig. 4 shows the inclusions in crystals grown by using lascas from Diamantina-I (Al = 30.1 ppm, Fe = 4.4 ppm) and Bicas (Al = 241 ppm, Fe = 8.2 ppm). At this moment, we do not know the mechanism of inclusion formation due to the influence of Al content. Conventionally, it has been reported

(5, 6) about the formation of Fe-rich compounds (Acmite) for inclusions. The chemical composition of these inclusion in the present research has not been analyzed yet. But we can suppose that there is probably an influence of Al-impurity in the formation of these inclusions.

3.4. Infrared absorption

The infrared absorption coefficient at 3500(1/cm) in the Z-region of synthetic quartz is represented as a function of Al content in lascas, as shown in Fig. 5. If we consider that fluctuations on the growth condition may happen, it can be observed for impurity content up to ~100 ppm of Al in lascas, that the

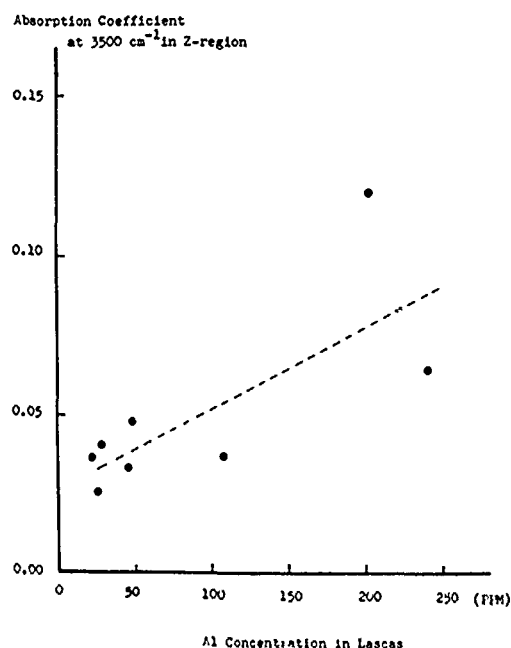


Fig. 5 Relationship between Al concentration and IR absorption in Z-region.

IR absorption coefficient will stay in a range below 0.05. Even though the increasing tendency of IR absorption coefficient in synthetic quartz with Al concentration in lascas has been observed, at present, we do not know the type of functional relation. Further experiments are in progress and they will be reported in the near future.

3.5. Ultrasonic evaluation

Ultrasonic attenuation measurements on synthetic quartz were performed by longitudinal pure mode on X- and Z-direction propagations in the frequency range from 10 to 90 MHz. The experiment was made at room temperature. Attenuation characteristics for measured and corrected values have been obtained for various synthetic quartz samples.

A typical attenuation plot is represented in Fig. 6. The extrapolation of the corrected or intrinsic attenuation curve for 5 MHz permits the evaluation of inverse-Q value through the expression

$$\alpha_0 \text{ (dB/cm)} = 8.686 (\pi f/V) Q^{-1},$$

where α_0 is the intrinsic attenuation of the elastic wave, f is the frequency and V is the propagation velocity of elastic wave in the specimen.

Table IV lists the inverse-Q values obtained for seven synthetic quartz samples and one natural quartz together with a comparison with the inverse-Q values obtained by other authors (7,8,9) in AT-cut plano-convex resona-

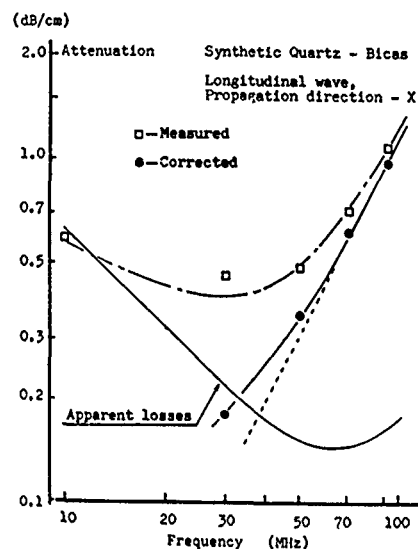


Fig. 6 Ultrasonic attenuations measured as a function of frequency in grown quartz

tors at 5 MHz (5th harmonics). Even though the order of magnitude of our results is in agreement with those found in literatures, the somewhat higher values may be caused by the elastic wave scattering on seed boundary of the synthetic quartz crystals used. Table V presents the normalized attenuation at 90 MHz for X-direction propagation (a_x) and Z-direction propagation (a_z) and the attenuation anisotropy (a_z/a_x).

Table IV Reduced Q^{-1} values at 5 MHz evaluated from ultrasonic attenuation in Y-bar quartz with comparison of Q^{-1} values of AT-cut vibrators.

Samples	Type of Vibrator	Propagation Direction	$Q^{-1} \times 10^6$ at 5 MHz	Ref.
Natural Quartz	-	X	1.34	7
Natural Quartz	-	X	0.31	8
Natural Quartz	AT	-	4.9	9
Synthetic Quartz - 812	AT	-	5.5	9
Synthetic Quartz - 688	AT	-	12.5	9
Synthetic Quartz - 644	AT	-	35.0	9
Natural Quartz	-	X	5.2	•
	-	Z	6.8	
Synthetic Quartz - Bicas	-	X	14.8	•
	-	Z	21.8	
Synthetic Quartz - Brumado	-	X	15.1	•
	-	Z	17.0	
Synthetic Quartz - Cabral (1)	-	X	11.1	•
	-	Z	23.3	
Synthetic Quartz - Cabral (2)	-	X	20.9	•
	-	Z	19.1	
Synthetic Quartz - Diamantina-I, (2nd)	-	X	16.3	•
	-	Z	21.7	
Synthetic Quartz - Diamantina-I, (3rd)	-	X	42.2	•
	-	Z	14.3	
Synthetic Quartz - Diamantina-I, (4th)	-	X	23.3	•
	-	Z	24.4	

• present works

We have tried to find the correlation between OH-bond concentration and ultrasonic attenuation in synthetic quartz crystals. However, at this moment, we do not find such a correlation.

Table V Ultrasonic attenuation anisotropies in Y-bar synthetic quartz crystals.

Samples	Normalized attenuations at 90 MHz (Longitudinal wave)		Attenuation anisotropy a_z/a_x
	a_x	a_z	
Natural Quartz	1.5	1.0*	0.67
Synthetic Quartz - Bicas	1.8	2.4	1.33
Synthetic Quartz - Brumado	1.8	2.5	1.39
Synthetic Quartz - Cabral(1)	1.8	2.6	1.44
Synthetic Quartz - Cabral(2)	3.3	2.1	0.64
Synthetic Quartz - Diamantina-I, (2nd)	2.6	3.2	1.23
Synthetic Quartz - Diamantina-I, (3rd)	5.0	2.1	0.42
Synthetic Quartz - Diamantina-I, (4th)	3.8	3.5	0.92

* 0.48 dB/cm

4. Discussion and Conclusion

The characterization results of synthetic quartz grown using lascas from various regions in particular with different Al content, show the following properties:

- the Al contents in the whole synthetic Y-bars decrease and depend on the original lascas, however, in the Z-regions, the differences of Al contents are not significant;
- there is an increase of s-regions in detriment to Z and +X regions and also an increase of strain field caused by growth striations as function of Al content in lascas;
- the solid inclusions increase with Al content in lascas;
- OH infrared absorption increases with Al content in lascas;
- the correlation between ultrasonic attenuation and OH concentration in synthetic quartz crystals has not been found.

At present, the purification mechanism of hydrothermal growth is not well known, but the experimental results reveal a fantastic high efficient effect for Al purification in the Z-region in the case of high Al content of the starting nutrient, for example, Bicas las-

cas with 241 ppm Al. Qualitatively, a similar result was observed in an experiment of growing quartz by doping 1000 ppm Al_2O_3 in weight of SiO_2 nutrient. (4) Experimental result shows that the growth rate of s-direction decreases by the presence of Al impurity. As a consequence, the area of s-region increases in detriment of Z-regions, being inconvenient for commercial use of Y-bar synthetic quartz.

With respect to the utilization of different grades of lascas, e.g., "2nd", "3rd" and "4th" from the same mine to grow crystals, the result shows that there is no significant difference in their quality for conventional resonator use. The grade difference is supposed to be related with the growth velocity of synthetic crystals because of the dissolving velocity depending on crack concentration of lascas. In this case, the characterization of lascas density will become an important parameter. The development of a research using lascas of lower grade for growing crystals is in progress as there is an increasing interest to apply the very abundant and low cost lascas, such as "4th" and "5th" for synthetic quartz mass production.

The relation between ultrasonic attenuation and OH concentration in synthetic quartz can not be well explained at this moment. On the other hand, Strakana and Savage show no correlation of infrared OH absorption and ultrasonic attenuation for fused quartz. (10) A study of correlation of ultrasonic attenuation and OH concentration in synthetic quartz will be continued in the future.

A study of elastic waves attenuation in specimens with different types and concentration of defects in synthetic quartz is in progress. The objective would be try to separate the effect of structural defects, such as inclusions, dislocations and boundaries, with the impurity concentrations.

Acknowledgement

The authors would like to thank Drs. Hideo Iwasaki and Fumiko Iwasaki, experts from Japan International Cooperation Agency for their valuable orientation and support throughout the development of this research. We acknowledge Claudio Silveira, Delson Torikai and Cristovão Farias for technical assistance. Financial supports by Fapesp and CNPq.

References

- (1) H. Iwasaki, F. Iwasaki, C.K. Suzuki, V.A. R. Oliveira, D.C.A. Hummel and A.H. Shino-hara, "Characterization of Brazilian Las-cas from Various Regions and Their use for Synthetic Quartz growth - Part I Las-cas Study, Proc. of 40th Ann. Frequ. Control Sympos. (1986)
- (2) B. Sawyer and D.R. Kinlock, "Cultured Quartz of Low Aluminum Content from Pro-duction Size Autoclaves", Proc. 38th Ann. Frequ. Control Sympos. (1984) 8
- (3) A.F. Armington and J.F. Balascio, "The Growth of High Quality Quartz in Comercial Autoclaves", Proc. 39th Ann. Frequ. Con-trol Sympos. (1985) 230
- (4) C.K. Suzuki, F. Iwasaki, and H. Iwasaki, "Influence of Aluminium in the Fabrication of Synthetic Quartz. Study by X-ray Topo-graphy and Goniometry", Presented at 29th Brazilian Ceramic Congress, Criciúma, SC (April, 1985) To be published on Revista Cerâmica.
- (5) J. Asahara, E. Yazaki and K.Kita, "Influ-ence on the Inclusions in Synthetic Quartz Crystal on the Electrical Characteristics of Quartz Crystal Resonator", Proc. 29th Ann. Frequ. Control Sympos. (1975) 211
- (6) R.L. Barns, P.E. Freeland, E.D. Kolb, R.A. Laudise and J.R. Patel, "Dislocation-Free and Low-Dislocation Quartz prepared by Hy-drothermal Crystallization", J. Crystal Growth 43 (1978) 676
- (7) J. Lamb, M. Redwood, and Z. Shteinshleifer "Absorption of compressional waves in Sol-ids from 100 to 1000 Mc/sec", Phys. Rev. Letters 3 (1959) 28
- (8) J. Lamb and J. Richter, "Anisotropic Acous-tic Attenuation with New Measurements for Quartz at Room Temperature", Proc. Roy. Soc., A293 (1966) 479
- (9) J.C. King, A.A. Ballman, and R.A. Laudise, "Improvement of Mechanical Q of Quartz by Addition of Impurities to the Growth Solu-tion", J. Phys. Chem. Solids 23 (1962) 1919
- (10) R.E. Strakana and H.T. Savage, "Ultraso-nic Relaxation Loss in OH-free SiO_2 ", J. Appl. Phys. 35 (1964) 1363

ON THE DISAGREEMENT IN THE ORDER OF MAGNITUDE
OF THE ELECTROELASTIC CONSTANTS OF ALPHA-QUARTZ

Carl K. Hruska

Piezoelectricity Research Laboratory
York University
Toronto, Ontario M3J 1P3, Canada

Summary

There exists a disquieting disagreement of almost three orders of magnitude in the published values of the electroelastic constants of alpha-quartz.

This work examines the differences among various sources in the collected experimental data and in their theoretical interpretation and concludes that by far the primary reason for the problem is improper algebraic and statistical treatment of some of the experimental data. The incorrect results are identified and the problem of disagreement is thus solved. The magnitude of the electroelastic tensor components $\partial c_{jk}/\partial E_i$ is of the order of 1 N/(V.m) and the standard errors in their determined values should be as small as several percent.

The values of the electroelastic tensor components obtained in 1982 by REIDER, KITTINGER and TICHY who used the transit time measurements are nearly in perfect agreement with the values calculated in 1977 by HRUSKA by means of the resonator method; the latter values are thereby independently and fully confirmed.

The work serves also as an illustration of the dangers of a mechanical use of data reduction methods and of statistics. The remarks made at the end are intended to focus the attention of the reader on this important but frequently unappreciated component of all processes leading to the determination of new material constants.

Introduction

The electroelastic constants in crystals are formally defined as the derivatives - with respect to the three individual components of the dc field E - of their twenty-one elastic constants. These derivatives are referenced to a chosen zero state which also includes $E = 0$. As such they represent the linear change in the elastic constants due to the electric field E .

The sixty-three individual derivatives thus obtained form together a fifth rank tensor of material constants which is called the electroelastic tensor. The tensor describes the so called electroelastic effect. It is one of the higher-order effects known to exist in quartz.

There exist two types of the elastic constants: the elastic compliances and the elastic stiffnesses. Consequently, one should generally speak about two electroelastic tensors, each related to one type of these elastic constants.

Well known thermodynamic relationships show an equivalence between the electroelastic tensor and the tensor describing the dependence of the piezoelectric constants on the mechanical stress or strain in

crystals. For this reason the electroelastic constants are also referred to as higher (the third) order piezoelectric constants.

The introduction of the electroelastic tensor of alpha-quartz as a measurable quantity by the author of this paper dates back to 1962 [1] shortly after it was demonstrated that the change in the frequency of quartz resonators caused by a biasing dc potential could not be explained merely by the converse piezoelectric effect [2].

The first attempt to determine the components of the electroelastic tensor related to the elastic compliances of quartz was made by HRUSKA in 1963 [3]. A similar attempt for the electroelastic tensor related to the elastic stiffnesses was made by HRUSKA and JANIK in 1968 [4]. KUSTERS in 1970 [5] was the first who tried to compute all the components of the latter tensor from one consistent set of measurements. BRENDL [6] should be credited with a first attempt in 1983 to replace the 'cumulative' component values of the entire tensor by the 'material' ones according to the theory formulated by BAUMHAUER and TIERSTEN [7].

Most of the work has been done on the electroelastic tensor related to the elastic stiffnesses. The reason for it is mainly the special attention which is generally paid to the quartz plates vibrating in thickness rather than to the rods vibrating in length and due to the presence of this tensor in the theory of wave propagation in quartz. Excellent work has been done there by the group of researchers around KITTINGER and TICHY. It is also this particular tensor which is the subject matter of this paper. From now on we will refer to it merely as the electroelastic tensor.

In spite of much serious effort over a period of almost two decades, there exist at present several sets of electroelastic tensor components which contradict each other in every possible respect: individual tensor components differ in sign and anywhere between two to three orders in magnitude. These disparities are discussed by KITTINGER, TICHY and FRIEDEL [8] who conclude that further study into the matter is necessary.

It is the aim of this paper to bring to the attention of the reader the singularly poor agreement among various authors, to clarify the reasons for the existing discrepancies, and to suggest the corrective measures which need to be taken.

Apart from the fundamental scientific interest a more accurate knowledge of the electroelastic tensor of quartz and other piezoelectric materials is required for optimization of the effect of the dc field on frequency and pulse transit times in potential applications such as fine frequency tuning and aging compensation in resonators, control of delay times in

SAW devices, and other applications based on the same principle.

Disagreement in the values of the electroelastic tensor components

Electroelastic tensor

The electroelastic tensor of alpha-quartz has eight independent tensor components. They are as follows:

(1) e_{111} , e_{113} , e_{114} , e_{122} , e_{124} , e_{134} , e_{144} , e_{315} .

They are generally defined as

$$e_{ijk} = (\delta c_{jk} / \delta E_i)_{E=0},$$

where c_{jk} ($j, k = 1, 2, \dots, 6$) are the elastic stiffnesses and E (E_i , $i=1, 2, 3$) is the electric field intensity. These derivatives are further understood to be taken at zero strain and constant entropy. While they are defined at constant entropy which corresponds to conditions of their measurement (vibrations), they are usually referred to the temperature of the quartz material at rest.

In quartz one finds altogether 23 non-zero components of the electroelastic tensor, all defined by means of the eight components listed above. The remaining components are zeros. The detailed diagram of the electroelastic tensor for quartz was originally given in [9]. Because of the notational changes that have become common since, the form of the tensor is restated here in TABLE I using matrix diagrams analogous to those used for presenting the elastic constants c_{jk} .

TABLE I

THE COMPONENTS OF THE QUARTZ ELECTROELASTIC TENSOR
RELATED TO THE ELASTIC STIFFNESSES
IN THE BASIC REFERENCE FRAME [13] DEFINED IN QUARTZ

components $e_{1jk} = (\delta c_{jk} / \delta E_1)_{E=0}$:

$$\begin{array}{cccccc} e_{111} & \frac{1}{2}(e_{111}+e_{122}) & e_{113} & e_{114} & 0 & 0 \\ & e_{122} & -e_{113} & e_{124} & 0 & 0 \\ & & 0 & e_{134} & 0 & 0 \\ & & & e_{144} & 0 & 0 \\ & & & & -e_{144} & \frac{1}{2}(e_{124}-e_{114}) \\ & & & & & -\frac{1}{2}(e_{111}+e_{122}) \end{array}$$

components $e_{2jk} = (\delta c_{jk} / \delta E_2)_{E=0}$:

$$\begin{array}{cccccc} 0 & 0 & 0 & 0 & -e_{124} & \frac{1}{2}(e_{111}+3e_{122}) \\ & 0 & 0 & 0 & -e_{114} & -\frac{1}{2}(3e_{111}+e_{122}) \\ & & 0 & 0 & -e_{134} & -e_{113} \\ & & & & e_{144} & \frac{1}{2}(e_{114}-e_{124}) \\ & & & & 0 & 0 \\ & & & & & 0 \end{array}$$

components $e_{3jk} = (\delta c_{jk} / \delta E_3)_{E=0}$:

$$\begin{array}{cccccc} 0 & 0 & 0 & 0 & e_{315} & 0 \\ & 0 & 0 & 0 & -e_{315} & 0 \\ & & 0 & 0 & 0 & 0 \\ & & & & 0 & -e_{315} \\ & & & & 0 & 0 \\ & & & & & 0 \end{array}$$

Disagreement in the published values

Illustrating the existing degree of disagreement among individual authors is a representative selection of results in TABLE II.

TABLE II

DISAGREEMENT AMONG SEVEN OF THE EIGHT COMPONENTS
OF THE ELECTROELASTIC TENSOR

independent tensor component	Kusters	Reider Kittinger Tichy	Brendel
e_{ijk}	[5] 1970	[10] 1982	[11] 1984
e_{113}	-49.05 *)	-0.49 $\pm 0.17^{**}$	-12.00 $\pm 1.2^{***}$
e_{114}	-28.81	0.39 ± 0.19	33.40 ± 6.1
e_{122}	-114.76	1.47 ± 0.14	3.05 ± 1.9
e_{124}	-85.49	1.13 ± 0.08	34.57 ± 7.2
e_{134}	2.12	1.35 ± 0.08	50.46 ± 8.2
e_{144}	57.77	0.15 ± 0.07	-0.604 ± 1.1
e_{315}	70.72	0.64 ± 0.08	24.90 ± 2.9

All values given in N/(V.m); *) tolerance 20%;
) standard error; *) type of error not specified.

The results [5] are published only in diagram form. Their actual unpublished values originate from [12].

Independently of their original published form, all the electroelastic tensor components given in TABLE II and further in this paper are stated for the right-hand quartz and the frame of reference according to [13].

The component values [5] and [11] are given for the temperature of 20 °C; the values [10] are presumed to be obtained for room temperature. However, this minor uncertainty is hardly a factor [14] when the problem under consideration is the order of magnitude of the respective quantities.

It is important to note that the disagreement in the

order of magnitude affects only seven out of the eight independent electroelastic tensor components: those shown in TABLE II. Indeed, it is befitting the arguments to be presented in this paper that it be just so.

Consistence of the experimental data

Resonator method

The results [5] and [11] in TABLE II are based on the measurement of the frequency-dc field dependence of quartz resonators also known as the polarizing effect. The effect consists in a frequency change Δf which occurs in a vibrating quartz resonator of frequency f if the body of the resonator is subjected to a dc biasing field ΔE .

The thickness vibrations of quartz plates were used here with the dc field applied in the plate thickness. The quantity of interest was the linear component of the relative frequency change defined as follows:

$$L = (1/f) \cdot (df/dE)_{E=0}$$

To calculate the values [5] (TABLE II) KUSTERS completed twelve observations of the quantity L , most of them repeated and confirmed a number of times on similar resonator samples [12]. The calculations performed by BRENDEL leading to [11] (TABLE II) were based on sixty-one independent values of L taken from HRUSKA [15] combined with data by KUSTERS [12] with the exclusion of a single outlier [16] identified in [17].

Transit time method

The values [10] in TABLE II were calculated from the changes ΔT in the transit time T of an ultrasound pulse travelling through a rectangular block of quartz. They were measured by means of the pulse-echo technique as a function of a biasing dc field ΔE . The measured quantity K [18] was the relative change in the transit time:

$$K = (1/T) \cdot (dT/dE)_{E=0}$$

To obtain their results [10] (TABLE II), REIDER, KITTINGER and TICHY used twenty-five independent experimental values of K .

There is no fundamental difference between the quantities L and K (except for the sign [19]); indeed, independently of the source [5,10,11], their typical order of magnitude is the same. Consequently, the experimental data could not create the problem in TABLE II.

Relationship between the experimental data and the electroelastic tensor

The relationship between the linear component of the polarizing effect L and the electroelastic tensor components as used in [5,11] is represented by the linear equation

$$(2) \quad L_i - C_i = a_{i1} \cdot e_{111} + a_{i2} \cdot e_{113} + a_{i3} \cdot e_{114} \\ + a_{i4} \cdot e_{122} + a_{i5} \cdot e_{124} + a_{i6} \cdot e_{134} \\ + a_{i7} \cdot e_{144} + a_{i8} \cdot e_{315}$$

where the index i indicates the dependence of the respective quantities on the crystallographic

orientation of the resonator, on its mode of vibration, and possibly also on the overtone number [20].

The simple fact that the principal dimensions and the density of the quartz resonators generally change in the dc field suggests that the linear component of the polarizing effect L_i cannot be accounted for exclusively by the electroelastic effect (e_{ijk}). In reality there exists a number of other phenomena adding to the value of L_i . Their total effect is represented by the correction term C_i which is duly subtracted from the experimental value L_i to isolate the contribution of the electroelastic constants e_{ijk} .

The coefficients a_{ij} ($j = 1, \dots, 8$) are relatively complicated functions of known material constants of quartz, the plate orientation and the mode of vibration. They are calculable from existing theory.

In the interpretation of the pulse-echo experiment as done in [10] the quantity L is replaced by K and equation (2) is adapted to the new experimental situation. However, its basic form as well as the order of magnitude of the quantities involved remain unchanged.

Equation (2) was formulated originally by HRUSKA in 1962 [1] and applied to the thickness modes of quartz plates in [21]; in 1973 it was improved by BAUMHAUER and TIERSTEN [7] who made it consistent with the state of the nonlinear theory of dielectrics. As a result of their work the values of e_{ijk} became calculable from equation (1) as true material quantities in the sense of their thermodynamic definition.

Formally the difference between [7] and [21] is reflected mainly in a different theoretical expression for the correction term C_i [6,21]. In both representations the values of C_i amount typically to about 20% of the size of L_i . As such, the practical effect of the differences between [7] and [21] is too small to account for the large disagreement in TABLE II.

Calculation of the electroelastic tensor

In order to determine the eight independent components of the electroelastic tensor a minimum of $m = 8$ equations of type (2) is needed. In order to alleviate the effect of the experimental errors in the measured quantity L_i as large a value of m as possible tends to be used. Thus an overdetermined system of linear equations is obtained with left hand sides affected by unavoidable experimental errors in L_i . This system can be recorded in a matrix form

(3)

$$\begin{pmatrix} L_1 - C_1 \\ L_2 - C_2 \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ L_m - C_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{17} & a_{18} \\ a_{21} & a_{22} & \dots & a_{27} & a_{28} \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ a_{m1} & a_{m2} & \dots & a_{m7} & a_{m8} \end{pmatrix} \cdot \begin{pmatrix} e_{111} \\ e_{113} \\ e_{114} \\ e_{122} \\ e_{124} \\ e_{134} \\ e_{144} \\ e_{315} \end{pmatrix}$$

where $m > 8$.

At this point a classical algebraic solution of the above system is no longer sought; it is rather tried to find a set of values e_{ijk} (1) (solution set (1)) that would satisfy (fit) the system best according to some chosen criteria. A convenient and frequently (though not exclusively) used method to find solution set (1) is the method of the least squares.

The form of the solution depends on the properties of the matrix of the elements a_{ij} , ($i = 1, 2, \dots, m$; $j = 1, 2, \dots, 8$), which will be referred to as matrix A. It is generally expected that one of these two distinct possibilities is encountered:

- (a) If the matrix A consists of columns (vectors) which are linearly independent, a complete solution set (1) of eight tensor components e_{ijk} can be obtained.
- (b) If there are some linear relationships among the columns, then the solution set will be incomplete; the amount of obtainable information will be reduced by the number of these relationships.

Insofar as the found solution set (1) fits the above system (3) well (and possibly even has a reasonable predictive power), it appears that its correctness is no longer doubted. Unfortunately, such has been the case with all three solution sets [5,10,11] listed in part in TABLE II; it is impossible for all of them simultaneously to be the sought-for electroelastic constants.

The 'weak' linear dependence and its effect on the solution set - an example

Properties of matrix A in general

The nature of the problem to be solved is numerical. It eludes the above simple algebraic considerations and requires that the properties of matrix A be discussed in some detail.

On the other hand it will suffice to concentrate on results [5] and [11] (TABLE II) i.e. on matrix A constructed for the interpretation of the data on the polarizing effect for the thickness vibrations of quartz plates with the dc field also applied in thickness.

The results by REIDER, KITTINGER and TICHY [10] (TABLE II) and the method of their calculation need not be discussed any further. They are not a part of this problem; later it will be re-confirmed that the order of magnitude of their results is correct.

Investigation of the properties of matrix A made here disclosed that - in a strictly mathematical sense - it consists of 8 linearly independent columns. However, some of the columns in matrix A could be expressed as linear combinations of other columns except for a residual which was much smaller than the original matrix elements. For lack of better terminology this case will be called the case of a 'weak' linear dependence. This phenomenon has a profound impact on solution set (1) of the problem.

Example: matrix A and solution set transformation

A simple example of the 'weak' linear dependence involving two arbitrarily selected columns in matrix A will be presented now to illustrate its effect on solution set (1).

Assume that it is possible to express the 8th column of matrix A in terms of the elements of its 7th column as

$$(4) \quad a_{i8} = a_{i7} + d_{i8},$$

where the quantities d_{i8} ($i = 1, 2, \dots, m$) are very small when compared to the typical size of elements a_{ij} ($d_{i8} \ll a_{ij}$), and at least some of the values d_{i8} are different from zero.* At the same time assume that the remaining 6 columns of the matrix are not interrelated in any other way similar to (4).

*In more precise language one should say: Assume that the regression of the last column on the remaining columns of matrix A yields the result $a_{i8} = a_{i7} + d_{i8}$, $i = 1, 2, \dots, m$, where d_{i8} represent the residuals which are much smaller than the typical size of matrix elements a_{ij} . - However, it is our intention to avoid the language of statistics as much as possible and to keep the argument at an illustrative common sense level.

If relation (4) is substituted into system (3), and, simultaneously, $Y_i = L_i - C_i$, $i = 1, 2, \dots, m$, is introduced as a space saving measure, system (3) can be recorded as follows:

$$(5) \quad \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{17} & a_{17}+d_{18} \\ a_{21} & a_{22} & \dots & a_{27} & a_{27}+d_{28} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{m7} & a_{m7}+d_{m8} \end{pmatrix} \begin{pmatrix} e_{111} \\ e_{113} \\ e_{114} \\ e_{122} \\ e_{124} \\ e_{134} \\ e_{144} \\ e_{315} \end{pmatrix} \quad *$$

Each individual equation of system (5), which now reads

$$(6) \quad Y_i = a_{i1} \cdot e_{111} + \dots + a_{i7} \cdot e_{144} + (a_{i7} + d_{i8}) \cdot e_{315},$$

can be rewritten to become

$$(7) \quad Y_i = a_{i1} \cdot e_{111} + \dots + a_{i7} \cdot (e_{144} + e_{315}) + d_{i8} \cdot e_{315},$$

where $i = 1, 2, \dots, m$.

Returning to the matrix notation, system (5) is now recorded in the following manner:

(8)

$$\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ Y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{17} & d_{18} \\ a_{21} & a_{22} & \dots & a_{27} & d_{28} \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ . & . & \dots & . & . \\ a_{m1} & a_{m2} & \dots & a_{m7} & d_{m8} \end{pmatrix} * \begin{pmatrix} e_{111} \\ e_{113} \\ e_{114} \\ e_{122} \\ e_{124} \\ e_{134} \\ e_{144}+e_{315} \\ e_{315} \end{pmatrix}$$

Matrix A of the original system (3) has assumed a new form that will be denoted A'.

The columns of matrix A' are still linearly independent, but its last column now consists of elements d_{18} which are much smaller than elements a_{ij} . At the same time one element of the original solution set (1), namely e_{144} , has turned into the element combination of $e_{144}+e_{315}$.

Properties of the solution set

Transformed solution set. Trying to find a fitting solution set for system (8), it appears possible to calculate all eight unknown quantities in (8) and ultimately the whole solution set (1). However, the unknown element e_{315} deserves our special attention.

Suppose that because of the smallness of coefficients d_{18} , the contribution of e_{315} to the left hand sides Y_i is not larger than the experimental errors in Y_i . Then, conversely, it is impossible to expect that this quantity can be calculated from the system with any degree of accuracy. If it is attempted to compute component e_{315} regardless of this fact, then the fitted value e_{315} will not pass a simple 'stability test'; namely, it can be shown that

(a) small (artificially introduced) variations in the values $Y_i = L_i - C_i$, which are well within the range of the experimental errors in L_i , will cause unexpectedly large changes in e_{315} amounting even to an order of magnitude or more;

(b) removing (or adding) one or several equations of type (7) from (to) system (8) will produce the same effect as in (a).

Furthermore, a properly calculated error in the computed element e_{315} will be unduly large and out of proportion with the experimental errors in L_i , thus reflecting the suspicious character of this result.

A more formal statistical analysis would show that element e_{315} is insignificant for the interpretation of the left hand sides Y_i and that it should be, together with the last column in matrix A', completely removed from system (8). The only reliable solution set obtainable from system (8) will consist of the top seven elements e_{ijk} , one being the element combination $e_{144}+e_{315}$.

Original solution set. Should the above case of the 'weak' linear dependence go unnoticed, it is very

unlikely that the transformation of the original system (3) into (8) would ever be made. For this reason, it is important to say what the impact of the 'weak' linear dependence on solution set (1) would be if it is sought by means of the original system (3).

Under these conditions the elements of solution set (1) associated with the 'weakly' interdependent columns, i.e. e_{144} and e_{315} , will fail the 'stability test' similar to the one described above. Again, the calculated errors in elements e_{144} and e_{315} would be unduly large. Frequently the values of e_{144} and e_{315} themselves would be completely out of the expected magnitude range.

None of the above should happen if e_{315} obtained by solving system (8) or e_{144} and e_{315} from system (3) are to be considered reliable. The fact that it does renders their values useless. They cannot be regarded as a measure of any material property nor do they themselves have any discernable physical meaning.

On the other hand, it can be shown that while elements e_{144} and e_{315} obtained by solving system (3) are each individually unreliable and unstable, their linear combination $e_{144}+e_{315}$, the form of which is dictated by the properties of matrix A, exhibits a remarkable stability. It is numerically close to its value obtained by the solution of system (8). The reason for this interesting fact is that systems (3) and (8) are - except for a different algebraic formulation - identical.

Also, the whole solution set (1) of system (3), including the unreliable values e_{144} and e_{315} , fits system (3) well and predicts new observations with accuracy (all this assuming that the quality of the experimental input and the theory are good).

Summary. The above can be summarized as follows:

(a) If a single case of the 'weak' linear dependence described above is present in the matrix A, then the total number of elements e_{ijk} and of their combinations that can be reliably determined from the system is no longer eight but only seven (i.e. their full number reduced by one).

(b) In order to obtain this maximum number of seven reliable values, the system must be solved in its transformed form (system (8)) rather than in the original one (system (3)).

(c) If the system is solved in its original form (3), then elements e_{ijk} of solution set (1) associated with columns of matrix A participating in the 'weak' linear relationship will not have stable and reliable values.

(d) The above mentioned properties of solution set (1) will not be evident when one is interested only in its interpretative and predictive potential. In fact, if the theory used to relate the experiment to the electroelastic tensor is a reasonable one, solution set (1) will fit the experimental data used for its determination and predict the magnitude of new independent observation with success.

Actual values of the electroelastic tensor components

Detected properties of matrix A and of the original solution set

In reality matrix A of system (3) contains two cases of the 'weak' linear dependence described in the

previous section. They are quite complicated and involve all columns of matrix A with the sole exception of the first one.

If the numerical properties of matrix A are unnoticed and solution set (1) is sought using system (3), then, consistently with the above example, seven of its elements e_{ijk} associated with the linearly interrelated columns will not be reliably determined. They will depend heavily on the experimental data used. No two independently determined solution sets of type (1) will likely be similar. Needless to say, they should not be regarded as the sought material constant of quartz. This is actually the case with the seven components e_{ijk} [5] and [11] in TABLE II.

Agreement in e_{111}

The first column of matrix A is not involved in the aforementioned linear relationships. Consequently, the values for e_{111} in [5] and [11] should agree reasonably well not only among themselves but also with the results of others. This can be seen from TABLE III. The percentages in TABLE III correspond to the errors stated by the individual authors.

TABLE III					
UNIVERSAL AGREEMENT OBTAINED FOR THE ELECTROELASTIC TENSOR COMPONENT e_{111}					
Hruska Janik	-2.30	20%	1968	[4]	
Kusters	-2.97	20%	1970	[5,12]	
Hruska	-2.55	8%	1971	[21]	
Graham	-2.64	2%	1972	[22]	
Hruska	-2.96	1%	1976	[15]	
Reider Kittinger Tichy	-2.61	3%	1982	[10]	
Brendel	-2.73	5%	1984	[11]	
All values given in N/(V.m); the type of errors is nonuniform and in part not specified.					

TABLE III supports our earlier remark that the existing disagreement among the values e_{ijk} occurs characteristically only with seven out of the eight independent components of the electroelastic tensor. The reason for it are the described numerical properties of matrix A.

Transformed solution set - tensor component combinations

Instead of the whole set of eight independent components e_{ijk} (1) one can determine reliably only a subset consisting of six values. Only one of these is a pure electroelastic tensor component; the remaining five values are their linear combinations. This

limitation is consistent with the example presented in the previous section of this paper.

The whole calculable subset is listed in TABLE IV. It was originally described by the author of this paper in 1976 [15]. The electroelastic component combinations are denoted c_{ijk} .

TABLE IV
CALCULABLE COMPONENTS AND COMPONENT COMBINATIONS OF THE ELECTROELASTIC TENSOR
pure component e_{111}
component combinations: $c_{113} = e_{113} - 0.42 e_{122} + 0.25 e_{134}$ $c_{114} = e_{114} - 0.26 e_{122} - 0.67 e_{134}$ $c_{124} = e_{124} - 0.76 e_{122} - 0.67 e_{134}$ $c_{144} = e_{144} + 0.50 e_{122}$ $c_{315} = e_{315} + 0.60 e_{122} - 0.51 e_{134}$

The above combinations c_{ijk} were obtained on the basis of a numerical analysis of matrix A. It can be shown that they are not unique. The numerical coefficients were derived for matrix A constructed according to [21].

Because of the complexity of the expressions defining matrix A, the reason for the existence of the particular combinations given in TABLE IV or the physical meaning of the numerical coefficients there have not been fully investigated so far. It can only be speculated that they may be somewhat dependent on the values of the material constants used to construct matrix A, on the finer details of the formulas used, on the quality of the computer program (e.g. personal preferences of the programmer when dealing with approximations) and also on the inevitable numerical computer clutter. At this point the reader is reminded again that we are dealing with a numerical problem.

The actual values of the calculable quantities listed in TABLE IV and obtained on the basis of 61 observations of the polarizing effect were originally given in [15]. For further reference, they are restated in TABLE V together with their computed standard errors.

Several properties of the quantities in TABLE V should be noted:

(a) Their order of magnitude is consistent with that of the other quantities in system (3). This is 10^{-12} m/V, 10^{-13} m/V and 10^{-12} m²/N for L_i , C_i and a_{ij} , respectively.

(b) Their standard errors are reasonably small and consistent with the standard errors in the measured quantities L_i which themselves are about 4% [15].

(c) As shown in [15] they are stable; they do not

TABLE V	
VALUES OBTAINED FOR THE CALCULABLE ELECTROELASTIC TENSOR COMPONENTS AND THEIR COMBINATIONS	
Hruska 197 [15]	
e_{111}	-2.96 ± 0.03
c_{113}	-0.68 ± 0.05
c_{114}	-0.90 ± 0.03
c_{124}	-0.88 ± 0.04
c_{144}	0.90 ± 0.02
c_{315}	1.01 ± 0.04
All values given in N/(V.m); also included are the standard errors.	

exhibit any strong dependence on the number and selection of the experimental data used for their calculation.

d) They successfully predict the magnitude of the linear component of the polarizing effect L_i in the entire primitive domain of quartz i.e. for all plate orientations. This has been illustrated in [17] and further verified on a large scale using over a hundred doubly-rotated plates [23]. A proper statistical analysis of the problem further indicates that, apart from the tensor component and combinations in TABLE V, no other information on the electroelastic tensor is needed for the purpose of making these predictions.

(e) The component combinations in TABLE V cannot be resolved by any additional experimental input from the measurement of the polarizing effect with quartz plates vibrating in thickness for the dc field applied in the same direction.

Stability of the component combinations

While the seven components [5] and [11] in TABLE II are unreliable, their combinations defined by the numerical properties of matrix A, should be stable and meaningful. TABLE VI shows how well these combinations of the electroelastic tensor components created by substitution of the values [5] and [11] from TABLE II into the relationships in TABLE IV agree among themselves and also with the directly calculated values from TABLE V.

In the light of the original differences among the individual electroelastic tensor components (TABLE II), up to almost three orders of magnitude, the fact that all above combinations have the expected magnitude is an impressive result. It provides solid evidence in support of the main idea of this paper.

In regard to smaller differences, such as those still present in TABLE VI, it should be pointed out that the three authors used in their work a physically different program or even a different data reduction method. Moreover, the experimental data used by KUSTERS contained an outlier spotted only at a later

TABLE VI			
COMPARISON OF THE COMBINATIONS OF THE ELECTROELASTIC TENSOR COMPONENTS CALCULABLE FROM THE POLARIZING EFFECT			
c_{ijk}	Hruska*	Brendel**	Kusters**
c_{113}	-0.68	-0.67	-0.32
c_{114}	-0.90	-1.21	-0.39
c_{124}	-0.88	-1.56	0.31
c_{144}	0.90	0.92	0.39
c_{315}	1.01	1.00	0.78
*Taken from TABLE V. **Based on the values [5] and [11] from TABLE II and calculated using the relationships in TABLE IV. All values given in N/(V.m).			

time [17]; BRENDL was attempting to determine the true 'material' constants whereas KUSTERS and HRUSKA computed only their 'cumulative' counterparts. Added to this is the effect of readings taken for overtone frequencies (which is yet to be analyzed [20]) and that of the electrostriction which was dealt with only marginally [15]. Under these conditions the differences existing in TABLE VI are considered justifiable.

Independent verification

A very important, independent verification of the values of the electroelastic tensor coefficients and their combinations as determined by HRUSKA in 1977 ([15], TABLE V) as well as of the validity of the previous arguments is provided by an agreement between these results and those obtained by REIDER, KITTINGER and TICHY in 1982 [10] (TABLES II AND III). It is shown in TABLE VII.

It should be emphasized that the compared sets of results in TABLE VII originate from the evaluation of experimental results obtained by two different methods: the resonator method and the transit time method. This independent verification should dispell any doubts concerning the agreement among the electroelastic tensor components very positively and persuasively.

Remarks concerning the use of statistics and of data reduction methods

The problem solved in this work was generated by unexpected and inconvenient numerical properties of a certain matrix which were not taken into account in two instances, being overlooked through an insufficient and improper use of statistics and of the data reduction methods. As it has not been an isolated case, it may be in place to make several remarks of caution, all of them relevant to our problem, but

TABLE VII		
NEARLY PERFECT AGREEMENT BETWEEN THE ELECTROELASTIC TENSOR COMPONENTS AND THEIR COMBINATIONS OBTAINED BY TWO INDEPENDENT METHODS		
e_{ijk} or c_{ijk}	Hruska* 1976	Reider** Kittinger Tichy 1982
e_{111}	-2.96	-2.61
c_{113}	-0.68	-0.77
c_{114}	-0.90	-0.90
c_{124}	-0.88	-0.89
c_{144}	0.90	0.89
c_{315}	1.01	0.83
*Taken from TABLE V. **Based on the values [10] from TABLES II and III and calculated using the relationships in TABLE IV. All values given in N/(V.m).		

thought to be also of more general interest.

It is not advisable to apply the formulas of statistics and data reduction methods while ignoring the associated statistical indicators and other available tools. A mechanical application of the formulas to fit a solution to an overdetermined system of equations produces only an untested result of doubtful value. It is only through the use of the techniques such as hypothesis testing and of the concepts such as significance probability that this result can be confirmed as a meaningful one.

The first and foremost statistical indicators are random errors. Statements concerning errors are frequently missing or they are not sufficiently informative. They will remain vague unless a proper method of calculation has been used and clearly identified.

Once calculated, the size of the errors must be accounted for. Failure to do so may mean that a possible inadequacy of the used theoretical model or a problem in the data reduction process will not be spotted.

It is not sufficient for a set of fitted values to be called (cumulative or true) material constants merely on the basis of a good fit or its capacity to produce good predictions. The problem discussed in this work is a point in case: in trying to fit a suitable set of eight electroelastic tensor components to the data on the polarizing effect, it is possible to construct an infinite number of solutions, all very satisfactory in terms of the two cited criteria. Not all of them can be the sought material constants. Two of the sets of the electroelastic constants examined in this paper are just two illustrations of this fact.

Statistical methods are usually applicable only if certain assumptions are satisfied. Undoubtedly, there exists a larger number of instances of the least square method being used than the number of cases of its applicability first being examined.

Conclusion

This work examined the differences, of several orders of magnitude, among the values of the electroelastic tensor components of quartz as published by various authors. Two sets of these values [5,11] have been found incorrect thus removing the principal disagreement.

The incorrect values were obtained owing to nontrivial numerical difficulties which have been either overlooked or not dealt with. Due to this fact the used standard data reduction procedures have not led to correct results.

The existence of the numerical problems has been indirectly implied earlier by the author of this work; unfortunately, their importance and general extent have not been, until now, fully appreciated. Due to their general character, the data on the polarizing effect obtained for the thickness modes of quartz plate resonators with the dc field applied in thickness can only yield six reliable electroelastic tensor components and combinations [15].

A most important result concerns the actual values of the electroelastic tensor components based on the polarizing effect measurements which were calculated by HRUSKA ([15], 1977). These values are fully verified by independent transit time experiments and calculations completed by REIDER, KITTINGER and TICHY ([10], 1982). A nearly perfect agreement has been attained there - not only in the order of magnitude, which was the main concern of this work, but also in the individual values.

There can be no doubt now that the of magnitude of the components of the electroelastic tensor is about 1 N/(V.m). Given the accuracy of the polarizing effect measurements, the standard errors in the electroelastic tensor components should not be greater than several percent. This is actually the case (TABLE V). Concerning the tensor component values based on the transit time experiment, their accuracy appears to be somewhat lower [10]; most likely this has to do with the nature of the used pulse-echo technique which is probably open to some improvement.

As a result of this work it is believed that the question of fundamental inconsistency in the values of the electroelastic tensor components no longer exists. Further refinements, such as the inclusion of the electrostriction, the transition from the 'cumulative' to the 'true material' quantities and possibly others, can take place meaningfully and effectively. It is understood that for the electroelastic constants based on the pulse-echo experiment the respective results will be published shortly [24].

Undoubtedly, a similar set of material constants will be produced also using the polarizing effect experiment. Given the high accuracy of the resonator method, which is based on the traditionally accurate frequency measurements, it would be a pity not to exploit this reliable source of information. It will be extremely interesting to see how well these new results will agree with those obtained by means of the transit time experiment. The good agreement between the 'cumulative' versions of these constants [10,15]

justifies some (cautious) optimism in this matter.

Acknowledgement

The author wishes to acknowledge the use he made of the unpublished results of measurements and values of the electroelastic constants which were kindly made available by Dr. J. A. KUSTERS as well as a most reassuring discussion with Dr. G. MONETTE. He is also obliged to Mr M. KUCERA, M.A.Sc., and to Miss P. HRUSKA for their assistance in the final stages of preparation of this paper. This work was supported by the National Sciences and Engineering Research Council of Canada (Grants G1146 and A4911).

References

- [1] Hruska K., "An attempt at a phenomenological interpretation of the influence of a polarizing field on piezoelectric resonators", Czech. J. Phys. B12 (1962), pp. 338-353.
- [2] Hruska K., "The influence of an electric field on the frequency of piezoelectric cuts", Czech. J. Phys. B11 (1961), pp. 150-152.
- [3] Hruska K., "Tensor of polarizing correction terms of quartz elastic coefficients", Czech. J. Phys. B13 (1963), pp. 307-308.
- [4] Hruska K., Janik L., "Change in elastic coefficients and moduli of α -quartz in an electric field", Czech. J. Phys. B18, (1968), pp. 112-116.
- [5] Kusters J. A., "The effect of static fields on the elastic constants of α -quartz", Proc. 24th Annual Frequency Control Symposium, U.S. Army Electronics Command, Ft. Monmouth, April 1970, pp. 46-54.
- [6] Brendel R., "Material nonlinear piezoelectric coefficients for quartz", J. Appl. Phys. 54(9), September 1983, pp. 5339-5346.
- [7] Baumhauer J. C., Tiersten H. F., "Nonlinear electroelastic equations for small fields superimposed on a bias", J. Acoust. Soc. Am. 54 (4), 1973, pp. 1017-1034.
- [8] Kittinger E., Tichy J., Friedel W., "Comments on 'Material nonlinear piezoelectric coefficients for quartz'", J. Appl. Phys. 56(9), November 1984, pp. 2584-2585.
- [9] Hruska K., Kazda V., "The polarizing tensor of the elastic coefficients and moduli for α -quartz", Czech. J. Phys. B18 (1968), pp. 500-503.
- [10] Reider G. A., Kittinger E., Tichy J., "Electroelastic effect in alpha quartz", J. Appl. Phys., 53(12), December 1982, pp. 8716-8721.
- [11] Brendel R., "Erratum: Material nonlinear piezoelectric coefficients for quartz", J. Appl. Phys. 55(2), January 1984, p. 608.
- [12] Kusters J. A., private communication.
- [13] IRE Standards on Piezoelectric Crystals, IEEE Standard 176; see also Proc. IRE, 37, December 1949, pp. 1378-1395.
- [14] Hruska C. K., "The dependence of the electroelastic constants of quartz on temperature", Proc. of the Quartz Crystal Conf., Electronics Industries Association, Kansas City, August 1986, in press.
- [15] Hruska C. K., "The electroelastic tensor and second-order phenomena in quasilinear interpretation of the polarizing effect with A-quartz plates", Proc. 31st Annual Frequency Control Symposium, U.S. Army Electronics Command, Ft. Monmouth, June 1977, pp. 159-170.
- [16] Brendel R., private communication.
- [17] Hruska C. K., "Polarizing effect with doubly rotated α -quartz plates vibrating in thickness", IEEE Transactions on Sonics and Ultrasonics, SU-25, November 1978, pp. 390-392.
- [18] Reider G. A., Kittinger, E., Tichy, J., "Correction procedure for measurements of the electroelastic effect in transverse fields", J. Acoust. Soc. Am. 71(5), May 1982, pp. 1283-1284.
- [19] Joshi S. G., "Surface acoustic wave propagation in a biasing electric field", J. Acoust. Soc. Am. 72(6), December 1982, pp. 1872-1878.
- [20] Hruska C. K., Kucera M., "Anharmonic character of the polarizing effect with the thickness modes of alpha-quartz plates", to appear in the IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control.
- [21] Hruska K., "Polarizing effect with piezoelectric plates and second order effects", IEEE Transactions on Sonics and Ultrasonics, SU-18, January 1971, pp. 1-7.
- [22] Graham R. A., "Strain dependence of longitudinal piezoelectric, elastic and dielectric constants of X-cut quartz", Phys. Rev. B6(12), December 1972, pp. 4779-4792.
- [23] Hruska C. K., unpublished results.
- [24] Tichy J., private communication.

INFRARED AND LASER SPECTROSCOPIC CHARACTERIZATION OF
ALUMINUM DEFECTS IN CULTURED QUARTZ

Herbert G. Lipson
Rome Air Development Center
Solid State Sciences Division
Hanscom AFB, MA 01731-5000

Summary

Optical absorption at He-Ne and argon laser frequencies has been used to map the distribution of aluminum-hole centers, Al-h^+ , in irradiated and vacuum swept cultured quartz. Al-OH^- and as-grown OH^- defect distributions were obtained for the same crystal region from infrared absorption scans. In this way all of the major quartz defect centers were evaluated over an entire crystal. The A_3 band associated with Al-h^+ has a strong peak centered at 435 nm and a weaker one at 633 nm. Argon laser lines at 488 and 514.5 nm are close to the wavelength of the stronger peak while the He-Ne line at 632.8 nm is nearly coincident with that of the weaker peak. Large variations in A_3 band absorption for irradiated and vacuum swept crystals correspond to visually observed changes in coloration. The laser method is sensitive and non-destructive and gives a more quantitative measurement than coloration.

Introduction

The principal defects involved in irradiation and sweeping (electrodifussion) of quartz are as-grown OH^- , a hydrogen-compensated center formed during growth by a hydrogen ion adjacent to an oxygen site, and those associated with aluminum substitutional in a silicon site. These defects and associated characterization techniques are reviewed in many publications.¹⁻⁷ Ionizing radiation dissociates the aluminum metal center Al-M^+ ($\text{M}=\text{Li}, \text{Na}$) and substitutional Al^{3+} is compensated with either hydrogen from OH^- in the crystal to form Al-OH^- , or a hole trapped at a non-bonding oxygen ion to form an aluminum-hole center $[\text{AlO}_4]^\bullet$, designated here as Al-h^+ . Sweeping also dissociates Al-M^+ , but in this case most of the alkalis are removed rather than being displaced to another site in the crystal. Air sweeping with external hydrogen available for Al^{3+} compensation produces only Al-OH^- . Vacuum sweeping can introduce either Al-OH^- or Al-h^+ , or both.

As-grown OH^- and aluminum concentrations cover a wide range in high quality quartz and can vary considerably across a single bar used for resonator fabrication. Non-destructive techniques are required to determine the concentration and distribution of these impurities after irradiation and sweeping in quartz selected for radiation-resistant resonators. We previously reported the application of low temperature infrared Fourier spectroscopy for scanning large crystals normal to the growth axis to sensitively monitor hydrogen-related absorption bands.^{8,9} From this technique initial OH^- and Al-OH^- distributions before and after irradiation and sweeping are determined. The aluminum profile across an as-received crystal can also be obtained from Al-OH^- absorption strengths after irradiation to saturation if sufficient hydrogen is available for compensation. In cases where hydrogen is depleted by irradiation, sweeping the sample in an air atmosphere provides the required hydrogen, and the Al-OH^- distribution should be proportional to the aluminum concentration.¹⁰

To completely characterize the aluminum-related centers a similar scanning method for determining Al-h^+ distribution is necessary. Al-h^+ as well as aluminum concentration can be measured from electron spin resonance (ESR)^{5,6}, but the technique requires small samples and is not easily applicable to non-destructive evaluation of distributions in large crystals. Optical absorption in irradiated quartz has been widely studied and visible coloration produced by irradiation is a well-known phenomenon. Mitchell and Paige¹¹ found good correlation between absorption bands at 450 and 620 nm and the aluminum content of quartz. Nassau and Prescott¹² showed a relationship between an absorption band with a peak at 2.9 eV (427 nm) and aluminum centers detected by ESR. Koumvakalis¹³ made a direct correlation between a visible absorption band with a peak at 2.5 eV (496 nm) and ESR signals from the Al-h^+ centers. More recently, low temperature optically detected magnetic resonance measurements at 1.6 K have shown that the A_3 band associated with Al-h^+ has a strong peak at 2.85 eV (436 nm) and a weaker one at 1.96 eV (633 nm) with respective half-widths of 1.0 and 0.55 eV for the π polarization (electric vector parallel to the optic c-axis).¹⁴ Weaker and broader bands are obtained for the σ polarization (electric vector perpendicular to the c-axis).

Conventional visible spectroscopy is not readily adaptable to scanning small regions of relatively thick samples. The small diameter, high intensity monochromatic beams of lasers are well-adapted to scanning crystals for impurity or defect absorption. Argon laser lines at 488.0 and 514.5 nm are close to the wavelength of the stronger A_3 peak and the He-Ne laser line at 632.8 nm is nearly coincident with that of the weaker peak. Thus either of these lasers should be useful for scanning quartz samples for A_3 band Al-h^+ absorption. The absorption obtained with argon lines will include a background whose main contribution is the A_2 band which peaks at 2.33 eV (532 nm), while the main background contribution with He-Ne line is the A_1 band which peaks at 1.77 eV (700 nm). In this paper optical absorption at laser frequencies is used to map the distribution of Al-h^+ centers in irradiated and swept quartz. From these absorption measurements and infrared scans over the same crystal regions an attempt is made to evaluate the contribution of all aluminum-related defect centers.

Experimental Procedures

The quartz samples used for this investigation were rectangular in shape, between 1 and 4 cm in size, with parallel x, y and z faces. Sample H29-14 was a SARP Premium-Q crystal grown with a lithium salt additive to the Na_2CO_3 mineralizer while E42-21, a High-Q crystal, was SARP grown by the same process, but without the lithium additive. QA-38 was grown at RADC from a NaOH nutrient with lithium additive. Seeds from z-growth sections were used for the SARP crystals while a seed from a +x-growth region was used for the RADC grown crystal. The aluminum concentrations for each sample were measured by electron spin resonance (ESR) from a single 2.5 x 2.5 x 8 mm section.¹⁵ Air sweeping, and vacuum atmosphere sweeping at 10^{-5} Torr, were performed

at RADC at 500°C with electric fields of 1000 to 1500 V/cm. Samples were irradiated in the RADC ^{60}Co source at doses up to 4 Mrad. The irradiation temperature was a few degrees above room temperature.

Infrared transmissions were measured between 3100 and 3700 cm^{-1} with the focused beam of a Nicolet 170SX Fourier spectrophotometer at 2 cm^{-1} resolution. Either a 1.5 or 3 mm diameter beam size was used. Overlapping scans taken at 1 mm or less were used for the measurements. Local variations were best determined with the 1.5 mm beam, but the better signal-to-noise ratio of the larger beam size was sometimes necessary to examine very low absorption regions. The sample, inside a Dewar cooled to 85 K, was scanned normal the z-axis with the unpolarized infrared beam aligned along the crystal x- or y-axis. More detailed description of the infrared measurement method, OH^- and Al-OH^- spectra, and the dependence of the absorption strength of these bands on crystal direction and polarization are given in our previous papers.^{8,9} For the purpose of this investigation where only relative distributions across the crystal is of interest, the peak absorptions of the strong narrow S_4 band at 3581 cm^{-1} and the e_2 band at 3366 cm^{-1} are selected to monitor relative change in as-grown OH^- and Al-OH^- , respectively. It should be pointed out that full strengths for these bands are not obtained with unpolarized radiation normal to the z-axis.

Optical transmission of a section of each sample was measured between 320 and 800 nm at room temperature on a Cary 14 spectrophotometer. The transmissions were obtained using unpolarized radiation with the beam covering a 5 x 10 mm area. These measurements gave the overall absorption of a relatively large section of the sample. The sample was scanned in more detail by laser measurements. The laser transmission measurements were made at room temperature with a randomly polarized 632.8 nm He-Ne laser. The laser power measured by a silicon detector was 0.15 mW and the unfocused beam size was 0.88 mm at $1/e^2$. The laser beam was propagated parallel to the crystal x- or y-axis and the sample was scanned along the z-axis. The laser beam was focused to about 0.2 mm to scan strongly varying crystal regions, but the unfocused beam, which had an intensity distribution close to that of the 1.5 mm infrared beam, was used for most of the measurements. Absorption levels at He-Ne and argon laser wavelengths were also compared. A vertically polarized argon laser with an unfocused beam size of 1.5 mm was operated in the simultaneous mode with all lines present at a reduced power level of 20 mW. The principal lines and their percentages of output power are 514.5 nm (37%) and 488.0 nm (32%). Additional weaker lines at 476.5 nm (12%), 496.5 nm (10%), 457.9 nm (4%) and 501.7 nm (5%) are also present. Polarization effects could not be fully evaluated due to the unavailability of both polarized and unpolarized He-Ne and argon lasers at the time of these measurements.

Experimental Results and Discussion

Room temperature absorption between 320 and 770 nm is shown in Figure 1 for sample E42-21 which was vacuum swept 8 days. The 5 x 10 mm section of this sample selected for measurement contained the darkest and correspondingly most absorbant regions. An aluminum concentration of 8 ppm was determined by ESR measurements made on another section of this sample. This spectrum is a composite of overlapping broad A_1 , A_2 , A_3 and B absorption bands.¹⁴ From Figure 1 it is

evident that this sample has relatively strong absorption at argon laser 488.0 and 514.5 nm and He-Ne laser 632.8 nm wavelengths.

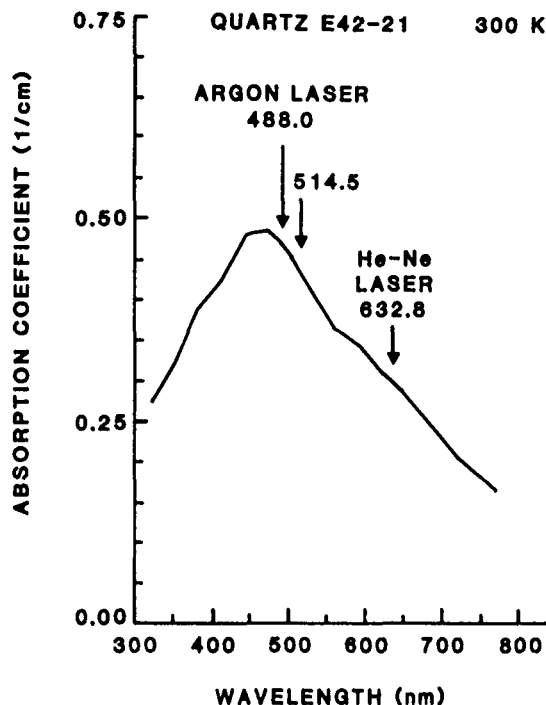


Figure 1: Room temperature optical absorption between 320 and 770 nm for sample E42-21 after vacuum sweeping 8 days. Principal argon laser lines at 488.0 nm and 514.5 nm and He-Ne line at 632.8 nm are indicated.

Figure 2 compares, (2a), variations in 632.8 nm He-Ne laser transmission observed along the z-, or sweeping axis, between the +z, anode, and z, cathode, faces with (2b) visible coloration for vacuum swept sample E42-21. Transmission scans were made for 15 positions along the x-axis with the beam directed along the y-axis and focused to about 0.2 mm. Good correspondence between major dips in transmission and the degree of darkening of the crystal is found. Some of the smaller transmission dips were not observed when the beam was defocused and are attributed to interference effects. The 320-770 nm absorption measurements of Figure 1 are for the sample region from 5-15 mm along the x-axis and 3-8 mm along the z-axis of Figure 2(b).

Figure 3(a) shows absorption variation along the z-axis for sample E42-21 determined from 632.8 nm transmission, for both focused and unfocused He-Ne laser beams. For x beam height 1.5 mm diameter infrared beam were made at the x-positions designated in mm. The coloration at these positions can be seen from Figure 2(b). The large differences between the absorption measured with focused and unfocused beams at the 13 and 14 mm x-positions are due to the different coverage of these strongly varying, dark colored, regions of the crystal. Large variations in 632.8 nm absorption found along both the z- and x-axis for this sample are an indication of a nonuniform Al-h^+ distribution.

Figure 3(c) shows that after 8 days of vacuum sweeping, total depletion of OH^- across the crystal is observed at the 7 mm position. For other positions the region of OH^- depletion extends from the anode to one half to two thirds the distance to the cathode. Figure 3(b) shows relatively high

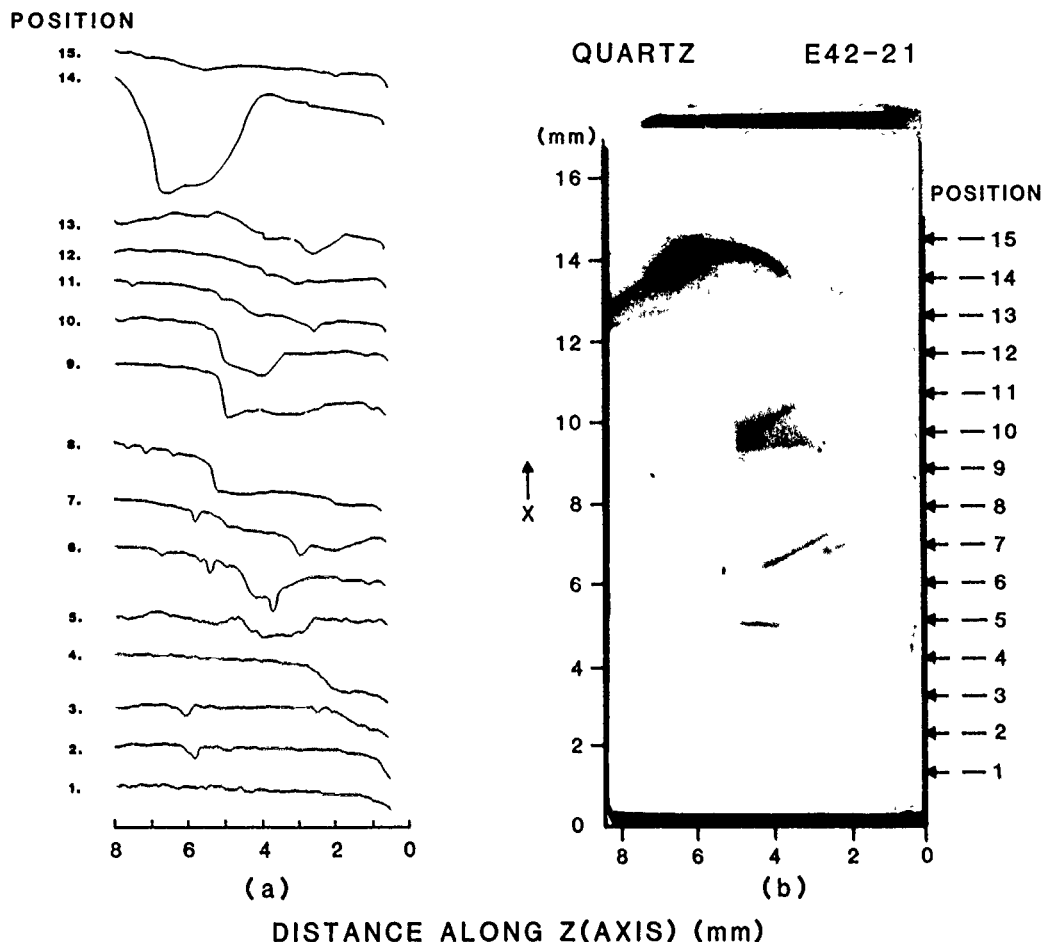


Figure 2: Comparison of (a) room temperature 632.8 nm He-Ne laser transmission scans along the z-axis for different x-positions and (b) coloration for sample E42-21 after vacuum sweeping 8 days.

Al-OH⁻ concentrations even for the OH⁻ depleted regions. Irradiation studies have led to the conclusion that an additional source of hydrogen other than as-grown OH⁻ is present in this material.⁹ This additional hydrogen source contributes to the formation of Al-OH⁻. Only for the totally OH⁻ depleted regions of the 7 mm and 8 mm positions is a nearly complementary behavior observed for Al-OH⁻ and Al-h⁺, with Al-h⁺ predominating when Al-OH⁻ is dissociated. Both forms of aluminum compensation are found in all other crystal sections. Regions of very high Al-OH⁻ and Al-h⁺ are evidence of high localized aluminum concentrations. Vacuum sweeping, if defined by the removal of all alkali and hydrogen impurities, has not been accomplished for this sample.

Figure 4 (a),(b) and (c) show the coloration after vacuum sweeping and the 632.8 nm, 3581 cm⁻¹ and 3366 cm⁻¹ absorption respectively, for three different regions of sample H29-14. The size and y-position of the infrared beam are indicated by the circles of Figure 4(a). This sample was first air swept by SARP and then vacuum swept by RADC in successive stages for a total of 52 days. After air sweeping both as-grown OH⁻ and Al-OH⁻ were present across the crystal. Average OH⁻ absorption values varied from 0.30 cm⁻¹ at the anode (+z end) to 0.68 cm⁻¹ at the cathode (-z end) while average Al-OH⁻ varied from 0.57 cm⁻¹ to 0.87 cm⁻¹. It is expected that the initial air sweeping removed all alkalis and substitutional aluminum is compensated with OH⁻. Figure 4(c) shows OH⁻ and Al-OH⁻ distributions after vacuum sweeping. OH⁻ is depleted or reduced and Al-OH⁻ is considerably

reduced between the anode and the center of the crystal. The higher Al-OH⁻ level near the anode for positions 2 and 3 may be due to sweeping-in of hydrogen impurities present on the crystal surface or electrode. High concentrations of Al-OH⁻ still remaining near the cathode are an indication that the vacuum sweeping was not complete.

The coloration of the sample and 632.8 nm absorption are in good agreement. The largest absorption values of Figure 4(b) correspond to the darkest regions of Figure 4(a), positions 1 and 2 close to the cathode. Al-h⁺ and Al-OH coexist in all regions of the sample with large concentrations near the cathode. A decrease in Al-h⁺ measured by 632.8 nm absorption is observed near the cathode for position 3 which has the largest concentration of Al-OH⁻ and remanent as-grown OH⁻.

Four sections of sample H29-14 were cut for ESR measurement. Sections 1 and 2 were close to infrared scan position 1, while sections 3 and 4 were close to infrared scan positions 3. Table I shows a comparison of 632.8 nm absorption and Al-h⁺ concentrations, as well as aluminum concentrations for these sections.

Reasonable consistency exists between Al-h⁺ determined from 632.8 nm absorption and from ESR measurements. Discrepancies may be attributed to the difficulty in matching the exact regions measured by both methods, or a more fundamental difference in the quantity determined by each technique. ESR measures Al-h⁺ absorption directly while absorption at 632.8 nm includes a sizeable background con-

QUARTZ E42-21

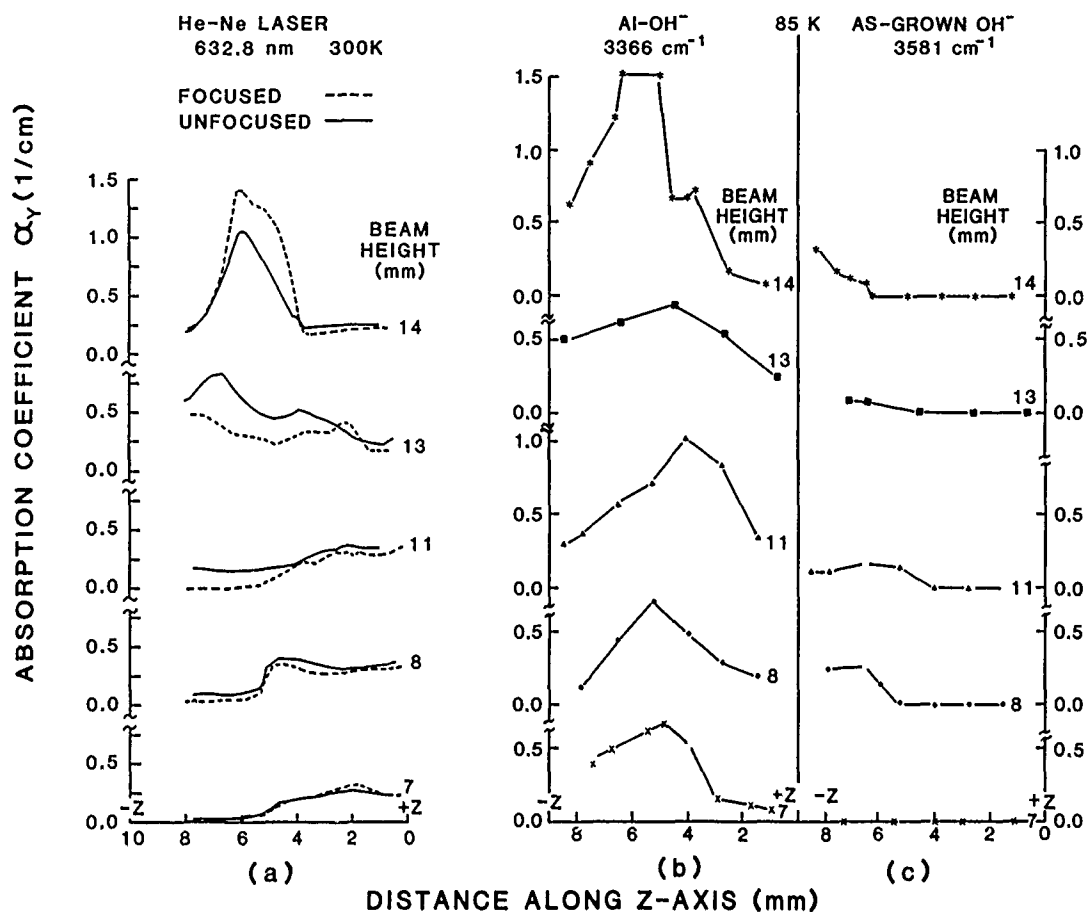


Figure 3: (a) 632.8 nm absorption (b) 3366 cm^{-1} and (c) 3581 cm^{-1} band peak absorption coefficient values as a function of position along the z-axis at different x-positions for sample E42-21 after vacuum sweeping 8 days.

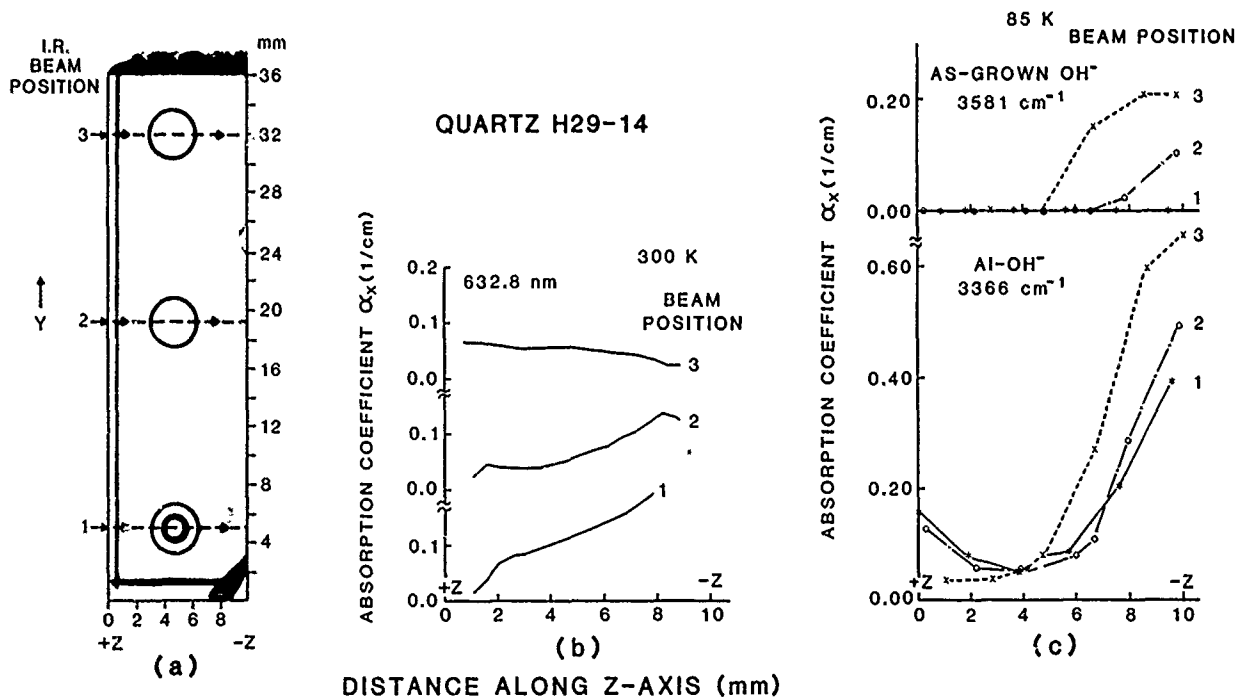


Figure 4: (a) Coloration, (b) 632.8nm absorption and (c) 3366 cm^{-1} and 3581 cm^{-1} band peak absorption coefficients values as a function of position along the z-axis for sample H29-14 after air sweeping and vacuum sweeping for a total of 52 days. The diameter and y-positions of the infrared beam are indicated in Figure 4a.

Table I

Sample H29-14, Air swept, vacuum swept 52 days. ESR determined Al-h⁺ and aluminum concentrations and 632.8 nm absorption for various sample sections.

Sample Section	Y Center Position (mm)	Z Center Position (mm)	Aluminum Concentration (ppm)	Al-h ⁺ (ppm)	632.8 nm Absorption (cm ⁻¹)
1	2.2	2.2	2.40	0.63	0.055
2	2.2	7.5	5.47	1.15	0.135
3	34.0	2.2	2.54	1.10	0.055
4	34.0	7.5	7.09	0.81	0.030

tribution from the A₁ band in addition to that of the A₃, Al-h⁺ band, which is not easily separated from these measurements at a single wavelength. Table I also indicates that Al-h⁺ to total aluminum concentration ratio for sections 3 and 4 varies between 43 and 11 percent. On the basis of the as-grown OH⁻ and Al-OH⁻ concentrations found for these sections, it is expected that a larger proportion of aluminum should be compensated with holes.

To evaluate the background contribution to total 632.8 nm absorption, the ratio of absorption at this wavelength to that measured in the vicinity of 500 nm is compared. The A₁ background band at the He-Ne laser wavelength of 632.8 nm and the A₂ background band at 500 nm, centered between the principal argon laser wavelengths, are nearly equal in strength.¹⁴ The 500/632.8 nm absorption ratio for the A₃ band is close to 1.8 at 1.6 K for π polarization.¹⁴ At room temperature the bands are broadened and a somewhat lower ratio of approximately 1.6 to 1.7 is expected. From spectroscopic measurements with unpolarized radiation covering the most absorbant regions of sample H29-14, the 500/632.8 nm absorption ratio was 1.33. This low value is evidence that the total absorption has a large background contribution. Sample E42-21, whose spectra is shown in Figure 1, has a stronger total absorption with a 500/632.8 nm ratio of 1.52, implying a smaller background.

Figure 5(a) compares the argon laser absorption between the +z and -z faces with the beam directed along the x-axis. Some differences are observed in the argon to He-Ne absorption ratio for the different y-height positions. Aside from these differences with y-positions, both lasers show the same variation along the z-axis. Figure 5(b) with the beam directed along the z- or c- axis. In this case the electric vectors of both lasers are normal to the c-axis (o-polarization) and the polarized argon laser has a larger contribution in the direction of maximum absorption than the randomly polarized He-Ne laser. The 500/632.8 nm ratio of about 2 observed for all of the y-height positions is attributed mostly to polarization effects. This method for estimating background absorption using either spectroscopic measurements, different laser wavelengths requires optical beams with the same polarization.

Z-growth sections of cultured quartz are customarily used for resonator manufacture, but the bordering -x-growth regions which can contain high OH⁻ or aluminum concentrations are occasionally accidentally included in the cut. The seed and z- and x-growth regions are usually delineated from the coloration produced by irradiation. 632.8 nm absorption and infrared scans are used here to examine the aluminum distribution and type of compensation for z- and -x-growth regions of low aluminum sample QA-38. An average of 0.5ppm Al was measured by ESR for the z-growth region of this crystal. After a 2 Mrad ⁶⁰Co irradiation 632.8nm scans were made at the x-height positions in mm shown in Figure 6(b), and infrared scans for OH⁻ and Al-OH⁻ determination were made at the positions indicated by the circles in Figure 6(a), which also designate the distance covered along the z-axis. The beam direction was along the x-axis for all measurements.

The outline of the seed and s-growth region can be seen from the dark colored areas of Figure 6(a). The z-growth region has a variation in shading. The darkened lines close to the edge of the photograph are from the back surface of the crystal. 632.8 nm absorption is an alternative method to visual coloration for observation of growth boundaries. The absorption measured at the various x-positions shown in Figure 6(b) give a good mapping of the coloration. The strongest absorption is in the seed and s-growth regions corresponding to the highest Al-h⁺ concentration. Absorption decreases in the -x-growth region with distance from the seed. At the 10.5 mm position the 632.8 nm absorption is 0.05 to 0.06 cm⁻¹, not much higher than average 0.04 cm⁻¹ level found in z-growth sections.

Figure 6(c) shows that the 2 Mrad ⁶⁰Co irradiation dose markedly reduces or depletes OH⁻ in the seed and z- growth regions. High concentrations of OH⁻ still remain in the -x-growth region, increasing and becoming more uniform with distance from the seed. From a comparison of Figure 6(c) and 6(d) it can be seen that the Al-OH⁻ profile follows the OH⁻ distribution across the crystal, except for the seed. The highest Al-OH⁻ concentration is at the border of the seed and -x-growth region. The highest Al-h⁺ concentration is also observed for this border, but Al-OH⁻ is strongest in the -x-growth region and Al-h⁺ in the s-growth region.

QUARTZ H29-14

----- ARGON LASER (457.9,476.5,488.0,496.5,501.7,514.5 nm LINES)

—— He-Ne LASER (632.8 nm LINE)

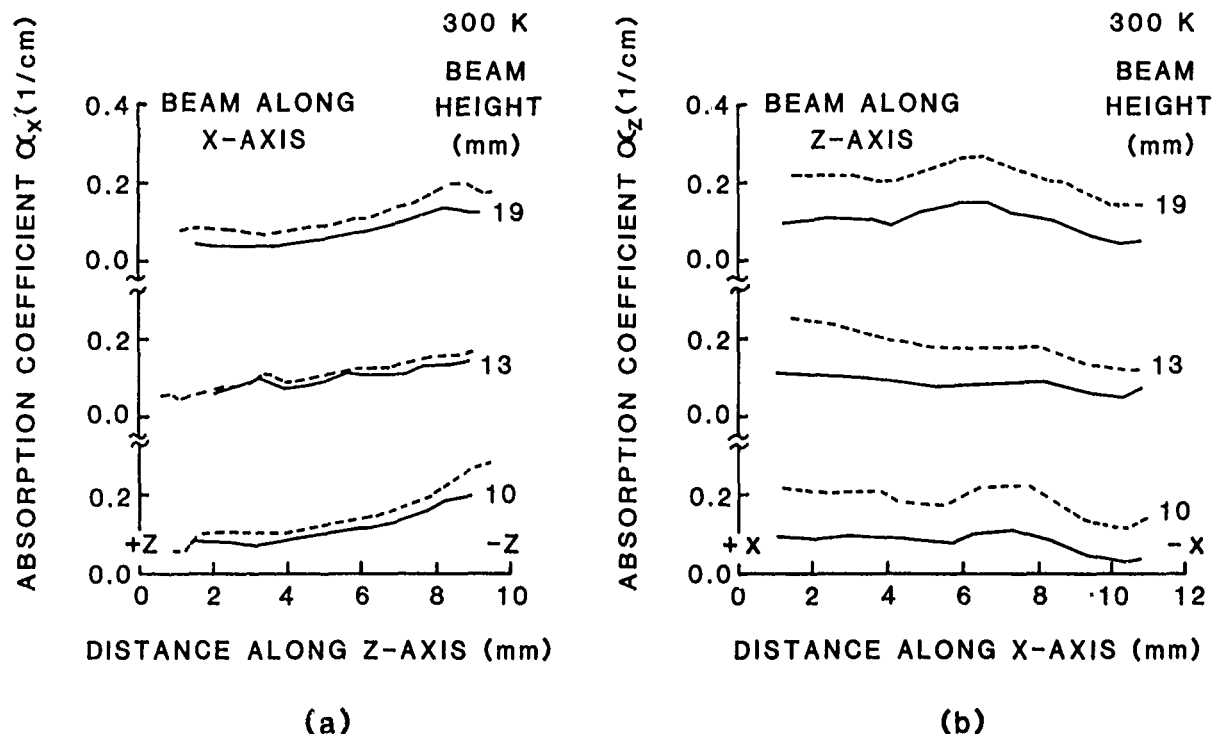


Figure 5: Comparison of argon and He-Ne laser absorption as a function of position for sample H29-14. (a) Absorption along the z-axis with the beam directed along the x-axis. (b) Absorption along the x-axis with the beam directed along the z-axis (optic axis).

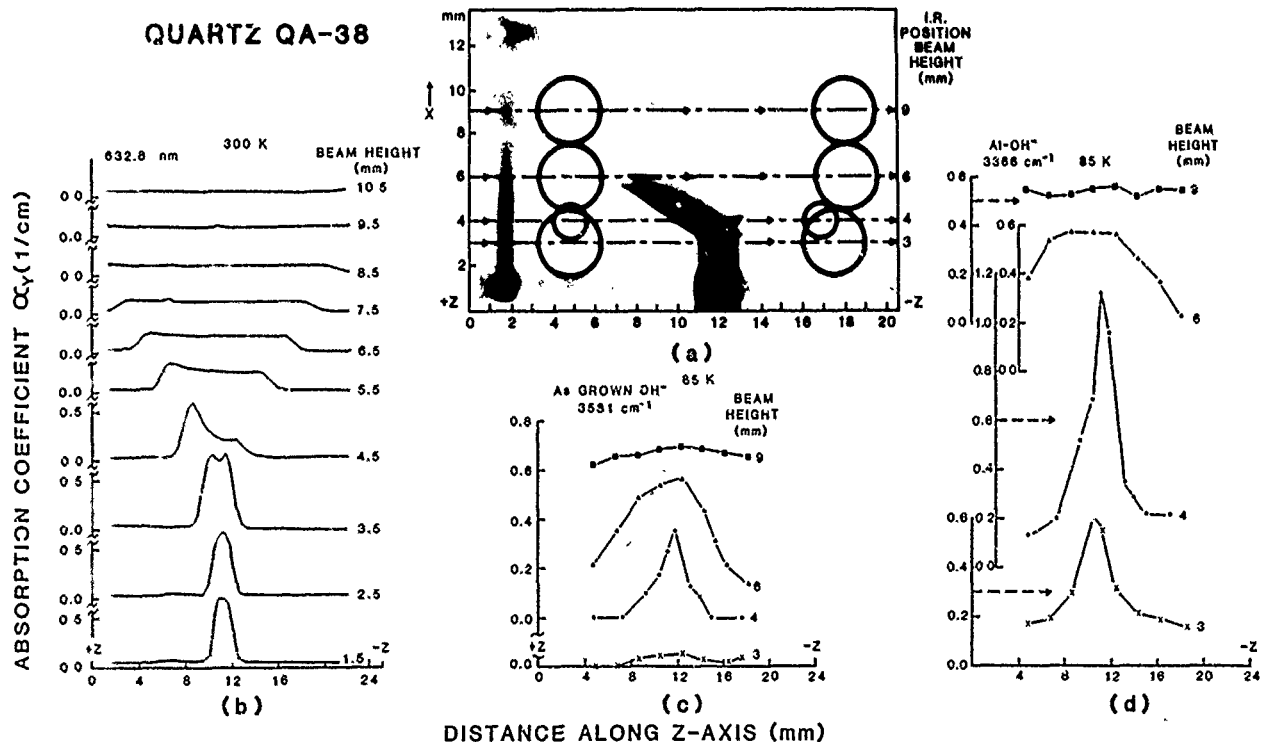


Figure 6: (a) Coloration, (b) 632.8 nm absorption, (c) 3366 cm^{-1} and (d) 3581 cm^{-1} band peak absorption coefficient values as a function of position along the z-axis for sample QA-38, showing -x- and z-growth regions after a 2 Mrad ^{60}Co irradiation. The diameter and x-positions of the infrared beam are indicated in Figure 6(a).

A variation in aluminum concentration of a z-growth section of QA-38 was determined from measurements of Al-OH^- after irradiation to saturation and compared with ESR results for sections taken approximately the same distance from the seed. Figure 7 shows the initial OH^- concentration of this sample and changes in OH^- and Al-OH^- produced by 1.6 and 3.2 Mrad ^{60}Co total dose irradiations and air sweeping for a period of 1 week. Almost no change in OH^- occurs from the second 1.6 Mrad irradiation while an increase in Al-OH^- is found toward the -z or seed end of the crystal. The sample was then swept in air for one week. The air sweeping restored the as-grown OH^- to nearly the initial distribution except for the small decrease close to the +z face sweeping anode. After air sweeping the Al-OH^- absorption level had the same distribution across the crystal. If the air swept Al-OH^- distribution is taken as the aluminum profile, a variation of 0.08 to 0.16 cm^{-1} is observed, a factor of 2 increase toward the seed over the sample length measured. The largest aluminum variation is toward the seed. ESR measurements gave aluminum values of 0.45 to 0.67 ppm, with the highest values toward the seed. Although this is a factor of 1.5 increase compared to 2 from the Al-OH^- determination, considerable error is introduced in exactly matching ESR sections and infrared beam positions, especially close to the seed where aluminum variation is large.

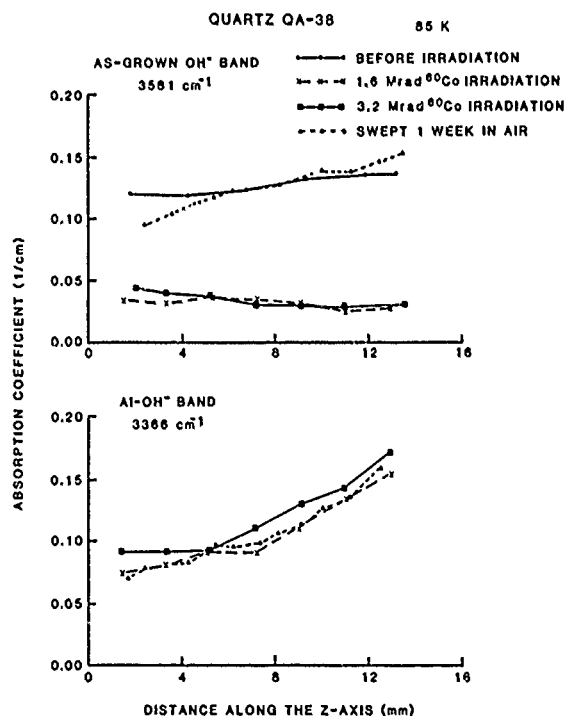


Figure 7: Peak absorption coefficient values of 3581 cm^{-1} and 3366 cm^{-1} bands as a function of position along the z-axis before and after irradiation and air sweeping for a z-growth section of sample QA-38.

Conclusions

He-Ne laser 632.8 nm absorption correlates well with coloration after irradiation or vacuum sweeping. The laser technique provides more quantitative information for mapping Al-h^+ distribution than coloration and is adaptable to computerized scanning and recording of data.

The combination of laser and infrared absorption scans provides information on the type of aluminum compensation present in each region of the crystal. Whether absorption at the He-Ne wavelength of 632.8 nm or at the argon wavelengths of 488.0 and 514.5 nm represents only the Al-h^+ distribution depends on the contribution of background absorption bands. The strength of background bands can be estimated from either optical spectra over a range of wavelengths or from measurements at both He-Ne and argon wavelengths. To evaluate the full contribution of the A_3 , aluminum-hole band, these measurements must be made with polarized lasers with the electric vector rotated parallel to the crystal c-axis.

The aluminum profile of a sample determined from Al-OH^- distribution after irradiation to saturation or air sweeping shows reasonable agreement with that obtained from ESR measurements.

Acknowledgements

The author would like to thank his co-workers at RADC for their contributions to this paper, A. Kahan and F. Euler for many helpful discussions, A. Armington and M.T. Harris for performing the sweeping experiments, J. Schott for the sample irradiation, and R. Andrews for assistance in performing measurements and data reduction. He would also like to express his appreciation to Professor L. Halliburton of Oklahoma State University for helpful discussions on interpreting electron spin resonance results.

References

1. A. Kats, Phillips Res. Reports **17**, 133 (1962).
2. D.M. Dodd and D.B. Fraser J. Phys. Chem. Solids **26**, 673 (1965).
3. D.B. Fraser, Fraser, *Physical Acoustics*, W. P. Mason, ed., New York: Academic Press, 1968, Vol 5, ch. 2, pp. 59-110.
4. R.N. Brown and A. Kahan, J. Phys. Chem. Solids **36**, 467 (1975).
5. M.E. Mrkes and L.E. Halliburton, J. Appl. Phys. **50**, 8172 (1979).
6. L.E. Halliburton, Koumvakalis, M.E. Mrkes and J.J. Martin, J. Appl. Phys. **52**, 3565 (1981).
7. H. Jain and A.S. Nowick, J. Appl. Phys. **53**, 477 (1982).
8. H.G. Lipson and A. Kahan, IEEE Trans. Nucl. Sci. NS-31, 1223 (1984).
9. H.G. Lipson and A. Kahan, J. Appl. Phys. **58**, 963 (1985).
10. A. Kahan and H.G. Lipson, Proceedings of the 39th Annual Frequency Control Symposium, 1985, [IEEE Document No. 85CH2186-5, IEEE 445 Hoes Lane, Piscataway, NJ 08854] p. 255.
11. E.W. J. Mitchell and E.G.S. Paige, Phil. Mag. **46**, 1353 (1955).
12. K. Nassau and B.E. Prescott, Phys. Stat. Solid (a) **29**, 659 (1975).
13. N. Koumvakalis, J. Appl. Phys. **51**, 5528, (1980).
14. B.K. Myer, F. Lohse, J.M. Spaeth and J.A. Weil, J. Phys. C: Solid State Phys. **17**, L31 (1984).
15. All ESR measurements for this investigation were performed by L. Halliburton at Oklahoma State University.

DEVELOPMENTAL RESULTS FOR THE PRODUCTION OF HIGH QUALITY QUARTZ

J. F. Balascio
Motorola, Inc.
Components Division
Carlisle, PA 17013
and

A. F. Armington
Electromagnetic Materials Technology Branch
Solid State Sciences Division
Rome Air Development Center
Hanscom AFB, MA 01731

Summary

The purpose of the investigation conducted during this phase of the work was to define the parameters of growth that would result in the development of a manufacturing method to produce cultured alpha quartz of such a quality that it would exhibit little or no frequency shift when exposed to continuous high radiation fields and to possess sufficiently low etch channel density so that it may be routinely fabricated into resonators capable of withstanding a high level of shock.

To meet the criterion of high frequency stability it was necessary to develop a method for the growth of high purity single crystal alpha quartz. The impurity ion that has the most deleterious effect upon the frequency stability is aluminum. More directly, the sodium ion associated with the aluminum cation effects the short term frequency stability of resonators. Since sodium is present in both common mineralizers employed to hydrothermally grow quartz, methods of purification had to be developed to reduce the aluminum level in the nutrient supply. Recrystallization of natural quartz and cultured quartz reduced the aluminum concentration below 3.0 ppm. Alpha quartz was then grown at reasonable rates with aluminum levels at or below 1.0 ppm in commercial size autoclaves. An estimation of the aluminum level and average Q could be obtained from the physical measurement of the crystal's X/Z ratio. A high degree of confidence was obtained under the controlled conditions employed for growth.

Low etch channel density alpha quartz was obtained by growing crystals upon +X(Z) seeds and by electrical sweeping. Typical channel densities were $<50\text{cm}^{-2}$ for crystals grown upon +X(Z) seeds and $<40\text{cm}^{-2}$ by electrical sweeping. Electrically swept quartz grown upon +X(Z) seeds typically had channel densities $\leq 1\text{cm}^{-2}$.

Average Q values equal to 2.9×10^6 were obtained on low ppm crystals grown in commercial size autoclaves.

Introduction

The cation impurities that are believed to have an effect on blank performance are the group one elements: lithium, sodium and potassium, as well as, aluminum and iron. The role of aluminum in alpha quartz is well documented and this impurity cation seems to be the most damaging to the properties of quartz. The roles that the other impurities play in alpha quartz are less well understood. Aluminum associated with alkali cations is the cause of frequency offsets since the alkali is easily disassociated from the aluminum by relatively low

energies. The employment of electrical sweeping results in the replacement of the group one impurities with hydrogen which is more strongly bound. This technique can essentially eliminate the active alkali impurities. The other impurity cation is iron. Data gathered to date indicate that the iron impurity ion does not contribute significantly to damage in alpha quartz. The specific goal with respect to impurities was to reduce the levels of each impurity to less than a part per million. A great deal of emphasis was placed on the aluminum concentration contained in the as-grown alpha quartz.

To meet these objectives, the investigations during this phase of the work centered upon the variables associated with the hydrothermal growth of alpha quartz. The variables studied were categorized as either process or component (Table 1).

PROCESS VARIABLES

1. VESSEL SIZE
2. NATURE OF MINERALIZER
3. THERMAL GRADIENT

COMPONENT VARIABLES

1. MINERALIZER PURITY
2. NUTRIENT PURITY
3. SEED PREPARATION
4. SEED TYPE

TABLE 1

The effect of changes in these variables and their effect upon the quality of the alpha quartz single crystals were assessed by employing EIA standard test procedures for Q and ρ (etch channel density).⁽¹⁾ Information on the impurity levels contained in the grown crystals were assessed by atomic absorption spectroscopy, ESR and inductively coupled mass spectroscopy. The bulk of the impurity data was obtained by atomic absorption spectroscopy.

During the course of this investigation, electrical sweeping runs were also conducted on quartz crystals grown in hydrothermal runs. When possible, mated half sections were prepared and one half section of the same crystal was electrically swept and its mate remained untreated. These mated sections

were then evaluated for their respective ppm levels and etch channel densities. The channel densities were usually measured on round "AT" blanks fabricated from these mated sections. The sweeping arrangement employed was presented elsewhere.

Results and Discussions

Process and Component Variables

Process Variables - Hydrothermal growth runs were conducted in unlined steel autoclaves with internal diameters of either four or ten inches. All runs conducted in the four inch I.D. vessel utilized lithium-doped sodium hydroxide as the mineralizer. In this paper these runs are designated by the letters ME preceding the growth run number. Ten hydrothermal runs were attempted in this vessel size. The remainder of the growth runs were completed in ten inch I.D. autoclaves. Those runs employing the same mineralizer as used in the four inch I.D. runs are designated by the letters MH and those employing the sodium carbonate mineralizer are designated with the letters MC preceding the run number. A total of thirteen hydrothermal growth runs employed NaOH as the mineralizer and seven with Na_2CO_3 in the ten inch I.D. vessels.

The thermal gradients ranged from one degree celsius to thirty-five degrees. Lower thermal gradients were employed for most of the growth runs in order to decrease the uptake of impurities into the crystal lattice and to increase the average Q value to levels greater than 2.5×10^6 .

Component Variables

Mineralizers - One of the goals of this program was to develop the ability to produce high purity, low etch channel density quartz from both the sodium hydroxide and sodium carbonate mineralizers. Typically prepared reagent grade mineralizer solutions were analyzed for their starting purity with respect to some of the impurities of interest. In addition, some sodium hydroxide solutions were analyzed for these same impurities after the growth run was completed. These data are shown in Table 2. The starting mineralizers were found to be of equivalent quality with neither possessing an unexpected impurity level nor a significant difference in levels between them. Analyses of the sodium hydroxide solutions, after growth, indicate higher levels for all impurities. ME-4 and MH-5 employed Cultured Quartz I as the nutrient supply and the aluminum level in both post-growth solutions were essentially equivalent. The primary source of aluminum is the nutrient supply. However, the increased concentrations of both potassium and iron have more complex origins and involve not only the nutrient but also the mineralizer and the vessel itself. However both of these cations are not readily substituted into the alpha quartz crystal lattice. Lithium was not analyzed since all of these solutions intentionally contained this cation.

Nutrient - During the conduction of these growth runs, nutrient supplies of different quality were employed. The preparation methods used to generate these alpha quartz nutrients were reported in a pre-previous paper contributed to this symposium. (2)

The majority of the growth runs conducted during this investigation employed Cultured Quartz I as the nutrient supply. This supply utilized the total cultured quartz as nutrient. Table 3 contains the ppm analysis for this nutrient supply by section. As was expected, the pure Z region sections of this nutrient were of the highest purity and contained the lowest

aluminum level. The section with the lowest purity was the +X sections. Two other trends were noted from this analysis. The sodium uptake decreased from the pure Z to the +X and the lithium increased from the pure Z to the +X. The indication from these data is that there is a relationship between the lithium and aluminum uptake. The higher sodium level in the pure Z region indicates that this direction allows the accommodation of larger cations compared to the other crystal sections. The average ppm values were calculated by using the relationships shown below the table. With an X/Z growth ratio of 0.600 the average weights per section can be calculated and an estimated percentage assigned for each section. The fractional equivalents per section for Cultured Quartz I nutrient are listed below. Employing these fractional equivalents the average ppm level for this nutrient was estimated.

COMPARISONS OF IMPURITY LEVELS IN
MINERALIZERS BEFORE AND AFTER GROWTH

MINER.	TIME.	RUN NO.	K (PPM BY WEIGHT)	AL	FE	NUTRIENT TYPE
Na_2CO_3	START	MC-1	0.7	<0.5	<0.5	NAT.
NaOH	START	ME-4	1.6	<0.5	<0.5	CUL. QIZ I
NaOH	END	ME-4	4.5	2.1	5.9	CUL. QIZ I
NaOH	START	MH-5	0.7	<0.5	<0.5	CUL. QIZ I
NaOH	END	MH-5	3.6	2.3	3.1	CUL. QIZ I

TABLE 2

IMPURITY LEVELS IN DIFFERENT SECTIONS OF CUL. QIZ I

SECTION	LI	NA	K	AL	FE	TOTAL
PURE Z	0.8	5.6	0.5	3.8	1.2	11.9
-X	3.1	2.8	0.5	7.3	1.2	14.9
+X	5.1	1.1	0.6	16.0	0.5	23.9
AVG.	2.0	4.3	0.5	6.9	1.0	14.7

$$X/Z = 0.6 = \text{GROWTH RATIO} = \frac{X(\text{CRYSTAL}) - X(\text{SEED})}{Z(\text{CRYSTAL}) - Z(\text{SEED})}$$

$$Z : -X : +X = 0.67 : 0.11 : 0.22$$

TABLE 3

Table 4 lists the relative distribution coefficient for aluminum as a function of the nutrient supply employed. A minimum of three independent determinations were performed to obtain the aluminum level in the various nutrient supplies. Crystals grown from these supplies

were then analyzed for their aluminum uptake and an average value calculated for the same. The number of samples used to obtain the average value are also shown. The aluminum level for both mineralizers was equal to 0.5 ppm and this concentration was then added to the average aluminum concentration determined for each nutrient supply. No other sources of aluminum were considered for the calculation of the relative aluminum distribution coefficient. As the purity of the nutrient supply increased, the relative distribution coefficient decreased. These data indicate that the Cultured Quartz III and the synthesized alpha quartz nutrient are nearly equivalent in their aluminum concentrations and the crystals grown from these supplies were also nearly equivalent with respect to this impurity.

RELATIVE DISTRIBUTION COEFFICIENT FOR ALUMINUM
AS A FUNCTION OF NUTRIENT SUPPLY

NUTRIENT	AL _{NUT}	SAMPLE NO.	AL _{XTAL}	K
NAT.	21.0	2	14.5	0.67
CUL. QTZ I	6.9	13	2.2	0.30
CUL. QTZ III	2.4	4	0.7	0.24
SYN.	2.3	2	0.6	0.21

$$K = C(\text{CRYSTAL}) / [C(\text{NUTRIENT}) + C(\text{MINER.})]$$

TABLE 4

The hydrothermal runs conducted in the four inch I.D. autoclave were primarily for the evaluation of the effect of nutrient purity on the resultant alpha quartz growth. Table 5 compares the aluminum uptake in the pure Z region of the as-grown quartz as a function of nutrient supply. In this table are also listed the average Q value measured on the same crystal and its X/Z growth ratio. As the purity of the nutrient supply increased the aluminum uptake decreased. The same cannot be said for the average Q value. The average values for Al ppm, Q and X/Z ratio are also shown in this table. These data were then used to determine if any linear correlations existed among the three variables. The second section of the table lists the results of these determinations. A linear correlation was found between the aluminum content and the measured X/Z growth ratio on the as-grown crystal. The degree of confidence for this relationship exceeds 95%. It was believed that a degree of confidence $\geq 90\%$ would be a reasonably good correlation. The column labeled "R" is the actually calculated linear correlation coefficient. The lack of correlation among the other pairs of variables may have been due to the fact that an inconsistent nutrient supply was used in these runs. This possible reason seems to be substantiated from the results obtained in the MH runs which were conducted utilizing Cultured Quartz I as the only nutrient supply employed.

Table 6 contains the same comparisons for the growth runs conducted in commercial size autoclaves utilizing the same mineralizer. In this group of

runs significant correlations were found among all of the variables examined. Utilizing these linear relationships, a number of limits were calculated. With respect to the first relationship listed, an X/Z growth ratio of 0.959 would result in 0 ppm aluminum in the crystal. A crystal with that growth ratio would have an average Q value of 3.9×10^6 . Utilizing this Q value in the third equation results in a 0.1 ppm level of aluminum in the crystal. These calculations show that the equations are consistent among themselves. It should be noted that these runs were conducted under a limited range of growth conditions and, therefore the reproducibility between runs conducted under the same thermal program was quite good. Table 7 contains the Q and growth rate data for all of the MH runs conducted. A large number of crystals from each growth run were measured for their respective average Q values. The crystals were selected from various tiers in each growth run and represent a reasonable average for each run. For the first six MH hydrothermal runs, the average Q value for all the crystals measured was 2.04×10^6 , whereas the average Q for the thirty crystals examined for the last five MH runs was 2.71×10^6 . The average Z-growth rates for both of these groups were calculated to be 0.70 and 0.47 mm/day respectively. Included in this table are the standard deviations for each set of measurements made. Relatively small deviations were found in each run with respect to Q and growth rate.

LINEAR CORRELATIONS IN EXPERIMENTAL GROWTH RUNS

RUN NO.	AL (PPM)	Q (10^6)	X / Z	NUTRIENT TYPE
ME-4	3.3	2.0	0.792	CUL. QTZ I
ME-7	0.6	2.6	0.859	SYN.
ME-10	1.7	2.3	0.824	CUL. QTZ II
ME-16	0.5	2.1	0.837	CUL. QTZ III
ME-18	0.5	2.5	0.912	CUL. QTZ III
ME-24	0.9	3.4	0.892	CUL. QTZ III
AVG.	1.3	2.5	0.853	

X	Y	R	DEG. CONFID.	SLOPE	INTERCEPT
X / Z	AL (PPM)	-0.77	>95%	-19.10	+17.5
X / Z	Q	+0.70	>85%	+8.00	-4.3
Q	AL (PPM)	-0.43	<75%	-0.90	+3.8

TABLE 5

Seed Preparation and Seed Type - In order to determine the effects, if any, of seed preparation procedures on the quality of the crystals grown upon these seeds, initial growth runs employed high Q (2.4×10^6) and low etch channel density seed material ($< 111 \text{ cm}^{-2}$). All seeds were 0°X and each possessed good x-ray orientation. In some of the growth runs, seeds cut from electrically swept quartz were employed. In addition to the use of 0°X pure Z seeds, some seed material was cut from the $+X$ region and x-ray oriented as a pure Z seed. The nomenclature employed in this paper for this type of seed is $+X(Z)$. This initial $+X(Z)$ seed, cut and oriented, is defined as the mother seed. The alpha quartz crystal grown upon the mother $+X(Z)$ seed is designated as the mother crystal. When mother crystals are cut into seeds,

they are designated as first generation +X(Z) and the crystals grown upon these seeds are first generation crystals.

LINEAR CORRELATIONS IN PRODUCTION GROWTH RUNS

RUN NO.	AL (PPM)	Q (10 ³)	X/Z	NUTRIENT TYPE
MH-6	3.3	1.9	0.598	CUL. QIZ
MH-11	2.7	1.9	0.551	CUL. QIZ
MH-11	2.7	1.9	0.551	CUL. QIZ
MH-13	2.8	1.9	0.529	CUL. QIZ
MH-13	1.7	2.0	0.563	CUL. QIZ
MH-17	4.1	1.9	0.566	CUL. QIZ
MH-17	3.3	2.0	0.570	CUL. QIZ
MH-20	1.3	2.5	0.710	CUL. QIZ
MH-26	1.8	2.3	0.758	CUL. QIZ
MH-31	1.4	2.4	0.688	CUL. QIZ
MH-32	2.0	2.7	0.785	CUL. QIZ
MH-32	1.2	2.8	0.764	CUL. QIZ
AVG.	2.3	2.3	0.634	

X	Y	R	DEG. CONFID.	SLOPE	INTERCEPT
X/Z	AL PPM	-0.70	99.01	-7.3	+7.0
X/Z	Q	+0.94	>99.91	+5.0	-0.9
Q	AL PPM	-0.65	>99.01	-1.3	+5.2

TABLE 6

COMPARISON OF AVERAGE Q AND AVERAGE Z-GROWTH RATE DATA FOR ALL MH HYDROTHERMAL RUNS

RUN NO.	CRYSTALS TESTED	AVG. Q (10 ³)	STD. DEV.	Z-RATE (MM/DAY)	STD. DEV.
MH-5	18	2.31	0.18	0.56	0.02
MH-6	12	1.94	0.11	0.80	0.01
MH-9	3	2.13	0.08	0.76	0.01
MH-11	6	1.84	0.05	0.79	0.02
MH-13	9	1.89	0.09	0.76	0.01
MH-17	9	1.87	0.06	0.71	0.00
MH-20	6	2.43	0.04	0.50	0.01
MH-26	6	2.82	0.29	0.47	0.02
MH-31	6	2.51	0.22	0.47	0.03
MH-32	7	2.89	0.09	0.47	0.01
MH-35	5	2.91	0.14	0.46	0.01

TABLE 7

All the seeds were inspected with a polariscope in order to determine whether or not strain was present. In the few instances when strain was detected in a seed, the location of the strained area was recorded and the seed was mounted in the seed rack so that it could be properly identified and the as-grown crystal separated upon conclusion of the hydrothermal run.

Five different seed preparation techniques were employed during the course of this investigation:

- 1) 7.0 molal NH_4HF_2 (ABF),
- 2) 48% HF,
- 3) 40% NH_4F + 48% HF (AFHF),
- 4) Hydrothermally polished (HP),
- 5) Untreated (N).

The letter designations will be employed throughout this paper when referring to a particular seed treatment technique.

The AFHF etchant was prepared in an equivolume ratio. ABF, HF and AFHF were used primarily, at room temperature with etching times not exceeding 60 minutes. AFHF was also used at elevated temperatures (75°C) and extended etching times (90 minutes). The purpose of the high temperature etching was to produce chemically polished transparent seeds. Transparent seeds were also produced in a number of autoclave runs employing either sodium carbonate or sodium hydroxide as the mineralizer. Typical etching times were 720 minutes at a temperature of

approximately 340°C. Table 8 summarizes the etching conditions employed for seed preparation. A comparison of some of the etch channel densities measured on "AT" slices from crystals in various hydrothermal runs with different seed preparation techniques is shown in Table 9. The data indicate that there appears to be a wide latitude in the selection of an etchant for seed preparation. The preparation techniques employed apparently are not the dominant factor in the growth of low etch channel density as-grown alpha quartz. Similar results were obtained from crystals grown upon 0°X pure Z seeds cut from electrically swept alpha quartz bars.

ETCHING CONDITIONS EMPLOYED

ETCHANT USED	ETCHING TIME (MIN.)	ETCHING TEMP. (C)
ABF	5-60	22
HF	5-10	22
AFHF	5-90	22:60:75
HP	720	340
N	----	----

TABLE 8

In addition to the studies performed on crystals grown upon pure Z 0°X seeds, crystals grown upon +X(Z) seeds were also evaluated with respect to their etch channel densities. Some of these seeds were obtained from alpha quartz crystals provided by RADCO/Hanscom Field. Growth runs were conducted utilizing three different generations of crystals grown upon +X(Z) seeds: mother, first and second generation seeds. Some of these seeds were located in growth runs conducted in both ten inch and four inch I.D. vessels utilizing sodium hydroxide as the mineralizer. Table 10 compares the Q values and etch channel densities of crystals grown upon +X(Z) seeds as a function of seed generation. These determinations were made on slices cut from the as-grown crystal.

In the two growth runs in which both mother and first generation +X(Z) seeds were used, the channel

densities were not significantly different between them. That is, no noticeable degeneration with respect to ρ was found. However, in run MH-35 a significant difference in ρ was found between crystals grown upon first and second generation seeds. The significant increase in channel density between alpha quartz crystals grown upon these seeds can be attributed to a small area in the second generation growth (7%) which resulted in a higher ρ in the as-grown crystal compared to the channel density in the rest of the material. The etch channel density covering 7% of the AT area was 1215cm^{-2} and the channel density for the remaining 93% of the area was approximately 58cm^{-2} . As can also be seen from these data, high Q values were attained with very low etch channel densities.

COMPARISON OF ETCH CHANNEL DENSITIES
WITH SEED PREPARATION TECHNIQUES

ETCHANT	ETCH TIME (MIN.)	ETCH TEMP.	CHANNEL DENSITY (/SQ. CM.)
ABF	5	22	135
ABF	10	22	139
ABF	15	22	220
ABF	30	22	403
ABF	60	22	211
HF	5	22	159
HF	10	22	133
AFHF	5	22	189
AFHF	10	22	155
AFHF	75	75	165
HP	720	340	102
N	0	---	185

TABLE 9

COMPARISONS OF Q AND P AMONG CRYSTALS
GROWN UPON X(XZ) SEEDS

GENER.:	MOTHER		FIRST		SECOND	
RUN NO.	Q (10 ⁶)	RHO (CM ⁻²)	Q (10 ⁶)	RHO (CM ⁻²)	Q (10 ⁶)	RHO (CM ⁻²)
MH-32	2.9	14	2.9	13	----	----
ME-34	3.0	5	2.8	5	----	----
MH-35	---	---	---	---	3.0	196

TABLE 10

Electrical Sweeping Results - For a number of the growth runs, as-grown crystals were selected for electrical sweeping. The sweeping arrangement employed was modeled after that described by Lipson, Euler and Armington.⁽³⁾ The electric field employed

for these runs ranged between 1.5 to 2.5 KV/cm. The soak temperature was maintained at 500°C and platinum electrodes were used. All of these electrical sweeping runs were conducted in air. The data gathered during these runs coincide with the published results of Brown, O'Connor and Armington.⁽⁴⁾ After all physical measurements were completed on the crystals selected for sweeping, each crystal was ground on four sides and the original seed removed. Each half section was x-ray oriented to within $\pm 5^\circ$ on the X and Z faces. One half of each crystal was electrically swept. A section of both halves was then removed for impurity analysis and the halves were then fabricated into material capable of producing either AT or SC blanks.

Table 11 compares the ppm data obtained by atomic absorption of samples from mated half sections of alpha quartz. From the data contained in this table, it definitely appears that the lithium cation is removed during the electrical sweeping operations. Sodium and potassium appear to present a mixed result with respect to removal. The difficulty in assessing whether these two cations are indeed removed during sweeping is due, in part, to the handling technique one might employ in sample preparation. The same can be said for the iron level. For runs MH-26, MC-28 and MH-31, plastic gloves were employed for all sample handling done in-house and the results appear to indicate that finger electrolytes may have contributed to alkali contamination. On the average, the data indicate that both lithium and sodium are removed during the sweeping operation. By the same reasoning it also appears that some of the aluminum ions are also removed during the sweeping operation. The average aluminum level in the electrically swept material is 2.0 ppm while the average aluminum content in the mated unswept alpha quartz is 2.1 ppm. Although these average values are within experimental error, three of the samples had aluminum level differences ≥ 0.3 ppm between the mated section. This result seems to suggest that some of the aluminum cations are in an interstitial site rather than a substitutional one. It was also noted that when substantial reductions occurred for both the lithium and sodium cations, after electrical sweeping, a significant reduction in the aluminum level also occurred.

COMPARISON OF IMPURITY LEVELS BETWEEN ELECTRICALLY
SWEEPED AND UNSWEEPED MATED SECTIONS

RUN NO.	SWEEP NO.	LI	NA	K	AL	FE	TOTAL
MC-22	E141	0.5	1.0	0.5	2.7	2.2	6.9
MC-22	----	0.7	1.4	0.5	3.2	0.5	6.3
ME-24	E143	0.5	2.0	1.3	1.0	0.5	5.3
ME-24	----	0.3	0.5	0.5	0.9	0.5	2.9
MH-26	E153	0.5	0.8	1.2	1.5	0.5	4.5
MH-26	----	0.5	2.0	0.9	1.8	1.4	6.6
MC-28	E233	0.0	0.2	0.1	3.0	0.1	3.4
MC-28	----	0.7	0.7	0.3	3.3	0.3	5.3
MH-31	E232	0.0	0.0	0.1	1.6	0.1	1.8
MH-31	----	0.3	0.0	0.1	1.4	0.1	1.9

TABLE 11

The other possible benefit of electrical sweeping of alpha quartz is the substantial reduction in the etch channel density. A number of these mated sections were fabricated into 15mm round AT blanks. Between three to five of these blanks from each mated section were examined for their respective etch channel densities. Table 12 compares these measured etch channel densities on a run by run basis. In all of these measurements, the total area of each blank was counted and the number was converted to channels per square centimeter. Typically, there was a tenfold reduction in etch channel density between the electrically swept blanks and their unswept counterparts. Only run MC-33 did not exhibit an order of magnitude reduction in etch channel densities. For most of the blanks fabricated from electrically swept alpha quartz, the average etch channel density was 23cm^{-2} . The final value achieved in the swept material may be a function of either the sweeping conditions employed or a limiting mechanism in the quartz crystal itself.

COMPARISON OF ETCH CHANNEL DENSITIES BETWEEN
MATED ALPHA QUARTZ SECTIONS

RUN NO.	SWEEP NO.	SWEPT (cm^{-2})	UNSWEPT (cm^{-2})
MH-20	E138	22 ± 4	380 ± 32
MC-22	E141	25 ± 9	253 ± 7
ME-24	E143	22 ± 3	247 ± 11
MH-26	E153	21 ± 11	410 ± 91
MH-32*	E259	1 ± 1	24 ± 17
MH-32	E258	39 ± 13	428 ± 21
MC-33	E261	85 ± 51	249 ± 36
AVG.		36 ± 23	328 ± 79

* = CRYSTAL GROWN UPON X(Z) SEED, NOT INCLUDED IN AVG.

TABLE 12

Sections of both swept and unswept lumbered bars that were fabricated into the round AT blanks were sent out for ppm analyses. Comparing the data in Tables 11 and 12 for runs MC-22, ME-24 and MH-26, the total impurity level does not appear to be related to the etch channel density measured on the unswept blanks. Inherently one would assume that the change in channel density should be related to the alkali migration. However, the data in these two tables presently do not indicate a significant correlation between channel density and alkali impurity content.

Acknowledgements

The authors would like to thank T. Walker, Motorola, Inc. and M. Harris; J. O'Connor; J. A. Horrigan and E. Cormier of RADC for their assistance during this investigation. The atomic absorption work was performed at Eagle Picher Research Laboratory, Miami, OK. This work was performed under RADC Contract F19628-82-C-0067.

References

- 1) EIA Standard. "Quartz Crystal Test Methods", EIA-477-1, May 1985 (pp. 1-8).

- 2) Armington, A. F. and Balascio, J. F., "The Growth of High Purity, Low Dislocation Quartz", Proc. 38th Ann. Freq. Sympos., May 1985 (pp. 3-7).
- 3) Lipson, H. G.; Euler, F. and Armington, A. F., "Low Temperature Infrared Absorption of Impurities in High Grade Quartz", Proc. 32nd Ann. Freq. Sympos., May 1978 (pp. 11-23).
- 4) Brown, R. N.; O'Connor, J. J. and Armington, A. F., "Sweeping and Q Measurements at Elevated Temperatures in Quartz", RADC-TR-79-175, May 1979 (pp. 1-30).

EXPERIMENTAL STUDY AND NUMERICAL SIMULATION OF QUARTZ CRYSTAL ETCHED FIGURES

C.R. TELLIER, N. VIALLE^{*} and J.L. VATERKOWSKILaboratoire de Chronométrie, Electronique et Piézoélectricité
E.N.S.M.M., Route de Grav, La Bouloie, 25030 Besançon Cedex, FranceSummary

A phenomenological description of the etching of quartz crystals in fluoride solutions is proposed to explain the dependence of the etch rate and of the final shape of etch figures on orientation. The dissolution is determined by firstly the formation of hydroxyl groups and secondly the substitution of OH ions by fluoride ions. Accordingly, the limiting concentration of the fluoride specie at the surface depends on orientation since the reaction is governed by the exposure of the O atoms at the surface. Secondly a numerical simulation of the dissolution based upon a two dimensional mathematical model derived from the Frank model is carried out. General analytical equations which characterize in terms of orientation the direction in which an element of the surface profile progresses in the bulk crystal are used to determine the shape of a moving surface profile at any time of etching. Experimental results on singly rotated quartz plates with $\theta \approx 37^\circ$, 49° , -19° and -39° are compared with the predictions of this graphical model. The final dissolution figures and the profilometry traces which are characteristic to the crystal orientation agree well with the results obtained from the two dimensional etching model showing how interesting is this graphical approach of the etching problem.

Introduction

Chemical etching of quartz plates is a technique¹⁻¹⁶ by means of which the construction of high frequency quartz resonators free of surface damage layers^{2,4,17,18} can be achieved. Among the many experiments on widely used AT and SC cuts^{1-9,11-15} satisfactory results^{1,6-9} were obtained even with pure fluoride solutions or HF etchants. But previously published results^{3,8,10,11,13-15,18,19} on the chemical etching of differently oriented quartz in ammonium bifluoride solutions have revealed that the final dissolution figures are characteristic to the orientation of the surface on which they are formed. Similarly the shape of a surface profile is directly correlated to the crystal orientation and to the direction of trace^{5,10,11,13-15,18,19}. For particular crystal orientations stable convex or concave background structures were observed^{10,11,13-15,18,19}. Irving²⁰ suggested that such background structures correspond to typical etch rate versus orientation plots.

The purpose of the present work is to investigate the possibility firstly, of proposing a simple etching model of the quartz crystal to explain the dependence of etch rate on orientation and secondly, of developing a two dimensional numerical model for the dissolution which permits to follow graphically the shape of the moving etched surface. Finally a comparison of experimental results with the prediction of numerical model is undertaken.

The chemical attack model

This section is concerned with a phenomenological description of the etching of quartz crystals in fluoride solutions with emphasis on mechanisms which cause the etch rate to be orientation dependent.

The reaction mechanism

In the atomic model of a perfect quartz crystal the Si atom is surrounded by four oxygen atoms forming a silica tetrahedron²¹. But the real quartz surface consists of Si atoms singly, doubly or triply bounded to the lattice (Figure 1)²¹. In water diluted systems the oxygen atom at the surface combines with a proton to form an hydroxyl group. Frequently fluoride solutions contain F^- and HF_2^- ions^{22,23}. Hence in a first step F^- ions may substitute hydroxyl groups^{21,24}. However it is generally considered^{21,24-26} that for surface atoms doubly bounded to the lattice this process of substitution involves the further adsorption of a bifluoride ion. This subsequent step in the mechanism is followed (Figure 1) by the transfer of the Si atom from one site to the other resulting to the detachment of an SiF_4 molecule.

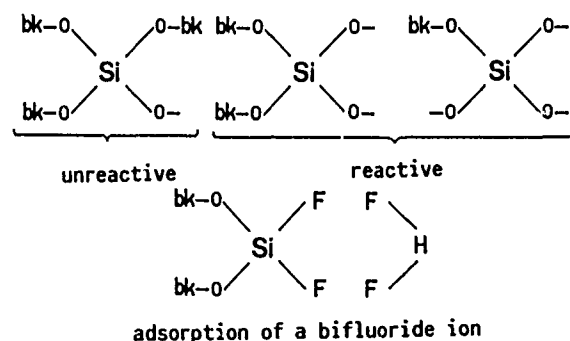


Figure 1 : The model for the chemical attack. The symbol bk is used for bulk.

Thus following the reaction mechanism previously proposed by Ernsberger a two reactions mechanism can describe the dissolution process of quartz.

1. Firstly, the breakdown, in aqueous solutions, of the bonds between a Si atom and the O atoms lying on the etching surface to form OH bonds.

^{*} Now with C.E.P.E. - ARGENTEUIL - FRANCE

2. Secondly and successively, the substitution of the OH^- ions by F^- ions and the absorption of an HF_2^- ion followed by the expulsion into the solution of the SiF_4 molecule. Since only surfaces on which two OH^- bonds are formed can react with the HF_2^- specie this second stage depends on the orientation of quartz surface.

The diffusion process

Let C be the concentration of etchant at large distances from the active surface (Figure 2), i.e. C is the initial concentration. Due to the reaction mechanism 2 the decay, $C - C(x)$, in etchant concentration at a distance x from the surface can be characterized by a lifetime τ defined by the relation

$$U = \frac{C(x) - C}{\tau} \quad (1)$$

where U can be identified with the rate of concentration decay.

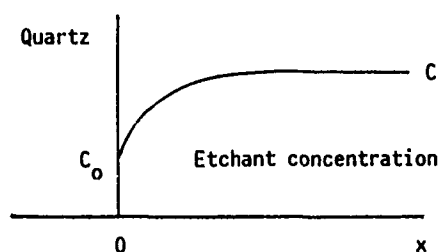


Figure 2 : The diffusion process.

As when the concentration is not uniform the reactant tends to diffuse toward the surface, i.e. from a region of high to a region of low concentration, a diffusion term must be taken into account. For the one dimensional case under the assumption that the diffusion rate is described by the well-known Fick equation²⁷ we have

$$\left. \frac{\partial C}{\partial t} \right]_{\text{diffusion}} = D \frac{\partial^2 C(x)}{\partial x^2} \quad (2)$$

where D is the diffusion coefficient.

The general form of the continuity equation is then

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C(x)}{\partial x^2} - U \quad (3)$$

This equation expresses the fact that the decrease near the surface in the concentration of the etchant is counterbalanced by the diffusion.

At steady state where $\partial C / \partial t = 0$ there is a concentration gradient near the surface and at the surface the concentration reaches a limiting value C_0 . Introducing a diffusion length, $L = \sqrt{DU}$, the differential equation now becomes

$$\frac{\partial^2 C(x)}{\partial x^2} = - \frac{C - C(x)}{L^2} \quad (4)$$

Since the boundary conditions are $C(x = 0) = C_0$ and $C(x \rightarrow \infty) = C$ the solution is given by

$$C(x) = (C_0 - C) \exp(-x/L) + C \quad (5)$$

Thus the reactant concentration decreases exponentially as it advances toward the surface to reach a limiting concentration C_0 at the surface. But this limiting concentration at the surface depends on orientation since the surface reaction 2 is governed by the exposure of the O atoms at the surface. As the dissolution rate is directly proportional to the concentration of the solvent at the surface it thus appears from this description that the orientation of the quartz crystal will play a significant role on the dissolution rate and consequently on the development of the typical etch figures.

The numerical simulation of the etching

The analytical equations

To provide a numerical simulation of the dissolution based upon a two dimensional mathematical model we can use analytical equations directly derived from the Frank model^{29,30}. The main features of this model can be characterized as follows. Let a singly rotated quartz plate of orientation, θ , be the reference surface and consider a local surface α degrees from the reference surface (Figure 3). This profile surface element $\Delta r(\alpha)$ of slope $p = \partial y / \partial x$ moves within the bulk with a velocity, v_y , normal to the reference surface such as

$$v_y = - \frac{\partial y}{\partial t} \quad (6)$$

Assuming that the position of an element of the local profile is a function of only two independent variables, namely the coordinate x and the etching time t , the continuity equation of Frank can be conveniently rewritten in the form

$$\frac{\partial v_y}{\partial x} + \frac{\partial p}{\partial t} = 0 \quad (7)$$

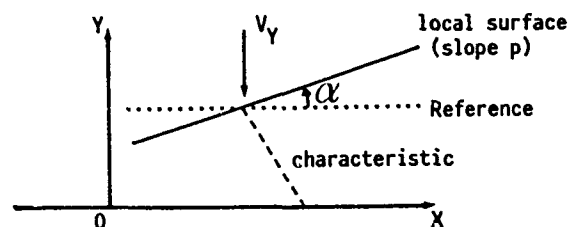


Figure 3 : The moving surface element in the Frank model.

Considering the model of the chemical attack presented in the above section we can restrict ourselves to the case where the velocity v_y depends only on the slope p i.e. on the local orientation of the quartz profile. Under this condition the equation (7) allows us to determine the linear trajectory, called a "characteristic"²⁹, of a moving profile element with slope p . A characteristic is thus a straight line of slope p_c

$$p_c = \frac{dy}{dx} = p - \frac{v_y}{\frac{\partial v}{\partial p}} \quad (8)$$

Thus after an etching time dt the initial element $\Delta r(y, x)$ of the local surface has progressed into the bulk in such way it is now located at $\Delta r_d(y_d, x_d)$ where

$$y_d = y + dy, \quad x_d = x + dx \quad (9)$$

with

$$dx = \frac{\partial v}{\partial p} dt \quad (10)$$

$$dy = dx \left[p - \frac{v_y}{\frac{\partial v}{\partial p}} \right] \quad (11)$$

Since the measured etch rate can be identified with the normal velocity V_N to a surface element the above equations characterize in terms of the slope, $p(\alpha)$, of the surface element and of the orientation dependent etch rate, $R(\alpha)$, the direction in which an element $\Delta r(\alpha)$ of the surface profile moves into the bulk crystal and allow us to achieve a numerical simulation of the etched surface profile.

The numerical simulation

The numerical simulation is accomplished in order to determine at any etching time the shape of a moving surface profile provided the variations of the dissolution rate, i.e. of the velocity V_N , with the orientation of surface profile elements were known. The two dimensional model derived in the above section is applied to surface profiles with initial randomized shape or with triangular shape.

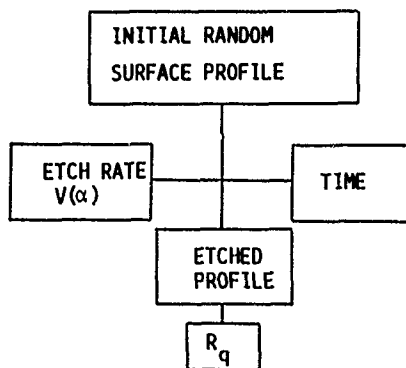


Figure 4 : The program for the numerical simulation.

The program (Figure 4) calculates the displacement of various points (x, y) of the initial surface profile and provides evaluation of the final coordinates (x_f, y_f) for successive etching times. Coordinates of potentially present final points

of the moving profile are also determined. These evaluations proceed with the selection of a typical etch rate versus orientation law. This program makes a numerical comparison of the x_f coordinates to distinct converging and diverging trajectories and to eliminate parts of the linear or the curved surface elements which disappear with prolonged etching. This program may be also used to estimate the initial and final r.m.s. surface roughnesses.

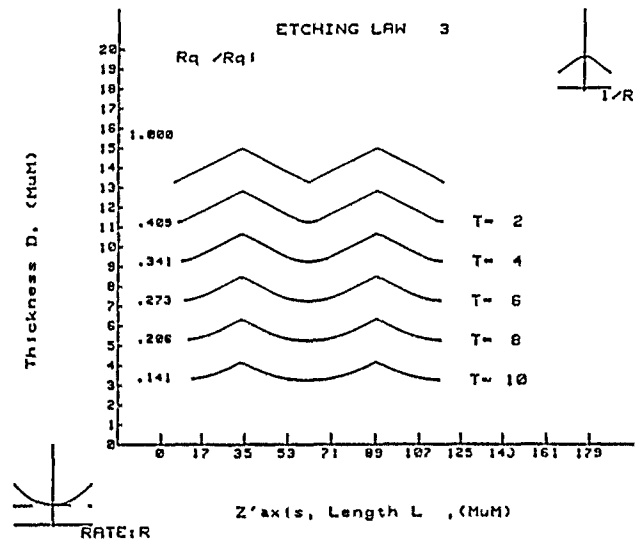


Figure 5 : The theoretical etched profiles. The case of an initial triangular profile.

Some typical results are displayed in figures 5 to 8. Figure 5 merits some comments. Effectively the mathematical model predicts that depending on the etch rate versus $p(\alpha)$ plot two successive surface elements may have converging or diverging trajectories. Converging trajectories result in stable intersections^{29,31,32} whereas two successive diverging trajectories corresponding to surface elements of orientations α_1 and α_2 give rise to a curved surface profile. The tangents to this curved profile correspond to faces whose orientations are between α_1 and α_2 . In other words only surface elements whose locus meet during etching do not change of shape on dissolution. Departures from this property mean the development of increasingly rounded profiles. This behavior is illustrated by Figure 5 where the etch rate, $R(\alpha)$, which is plotted against the slope, $p(\alpha)$, presents a minimum for the reference surface, i.e. for $p = 0$. We observe that the "convex" intersections are stable in accord with Irving²⁰ which concluded that a convex intersection between two crystal planes is stable provided there is no plane between them with higher etch rate. In contrast this typical etch rate vs slope plot results in the development of curved concave intersections which enlarge with prolonged etching.

If we regard the case where we substitute a randomized initial surface profile to the triangular profile it appears that a final concave background structure develops (Figure 6) as expected for an etching process which is primarily governed by the crystal orientation. Besides the effect of etch rate orientation relationship is depicted in figures 6 to 8 where different etch vs orientation

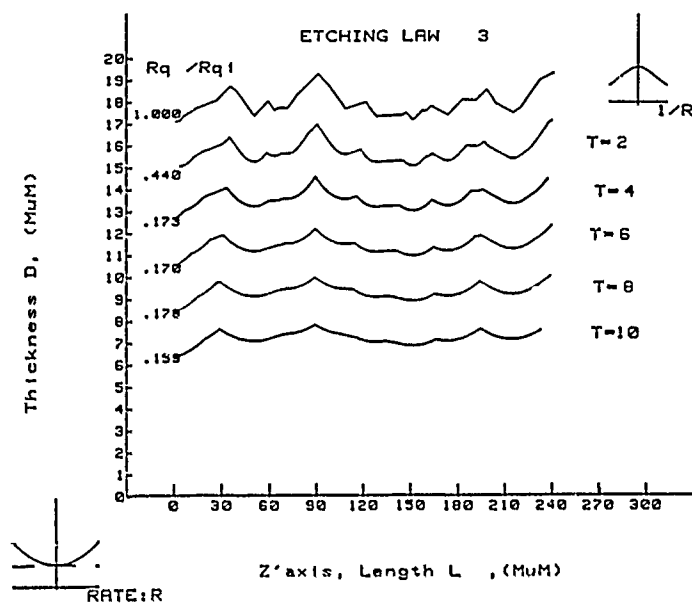


Figure 6 : The theoretical etched profiles. The case of an initial random profile and of a $R[p(\alpha)]$ relationship 3. The R versus p plot is shown in the left lower corner of the figure.

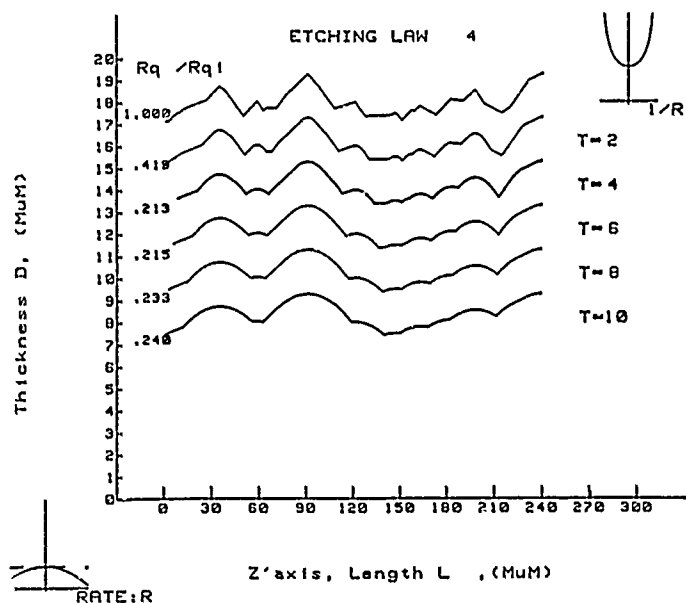


Figure 7 : The theoretical etched profiles. The case of an initial random profile and of a $R[p(\alpha)]$ relationship 4.

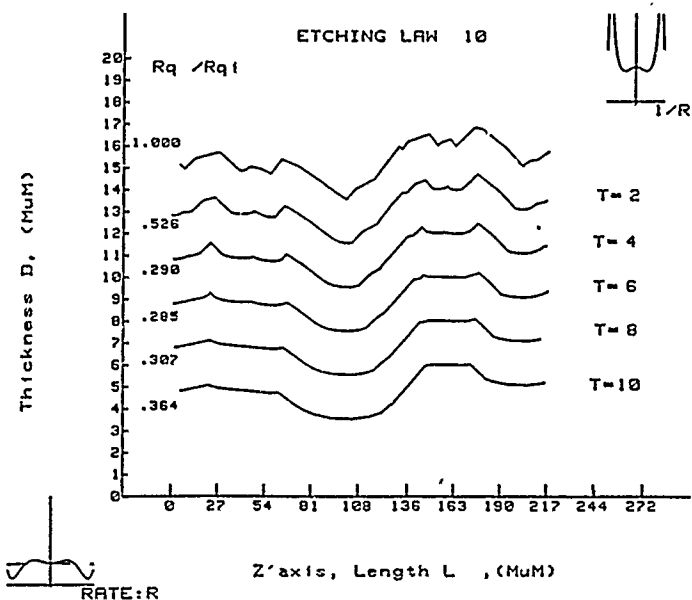


Figure 8 : The theoretical etched profiles. The case of an initial random profile and of a more sophisticated $R[p(\alpha)]$ relationship 10.

plots produce different final surface profile features.

Experimental study

Experimental procedure

Singly rotated quartz plates with various orientations (table 1) were cut from similar synthetic quartz bars. Successive etchs were performed in a concentrated ammonium bifluoride solution. After each etching step the plates were rinsed in distilled water and dried. The decrement in thickness was then evaluated from resonance frequency measurements.

=====					
: Quartz plates	:	Orientation, θ (degree)	:		
-----		-----		-----	
: AT - 37	:	37.5	:		
: AT - 49	:	49	:		
: AT - 19	:	- 19	:		
: BT - 39	:	- 39	:		
=====					

Table 1 : The various singly rotated quartz plates.

Prior to etch the different quartz plates were lapped with a 5 μm abrasive powder. At any time the topography of the quartz surfaces was studied by means of two different procedures. Firstly the geometry of each plate was examined by scanning electron microscopy along an observation angle of zero degree. Then the surface profiles were characterized by using a microprocessor based profilometer. The traces were made along two specified rectangular directions of the surface of various quartz plates, namely along the crystallographic X axis and along the Z' axis which lies within θ degrees of the Z axis.

Experimental results

Changes in the profilometry traces at various stages of etching are displayed in figures 9 to 12. As previously observed for singly or doubly rotated quartz plates the morphology of the surface profiles depends on the direction of trace. Profilometry traces on BT-19 and BT-39 surfaces are particularly indicative of directional effects.

All the figures reflect marked orientation effects. In view of the results we can classify the surface profiles in four groups. We distinguish without ambiguity the stable convex background structure corresponding to the final Z' profile of BT-19 and AT-49 quartz plates. The opposite stable type of background structure, concave, is depicted in X and Z' traces of the deeply etched AT-37 plates. The Z' trace of the BT-39 quartz plate exhibits a somewhat complicated final shape : concave bottomed pits are revealed with prolonged etching but the nearly concave section shows planes (denoted α_1 and α_2 in Figure 12) whose orientation α remains closely unaffected with repeated etchings. Finally we observe that successive etchings may also produce a relatively flat surface profile elongated along a particular direction : the X profile of BT-19 plates shows these features. At this point we note that a prolonged etching causes in general an enlargement of the surface profile. This enlargement is particularly marked in the case of AT-37 plate (X and Z' directions) and of BT-19 plate (X direction only).

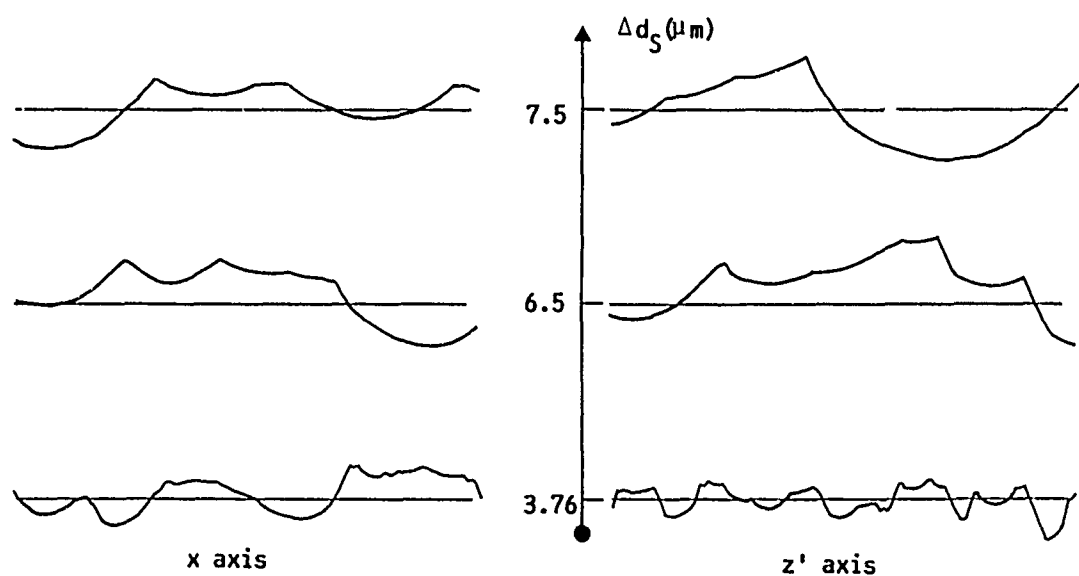


Figure 9 : Changes in the surface profilometry traces with the depth, Δd_s , of etch of an AT-37 quartz plate.

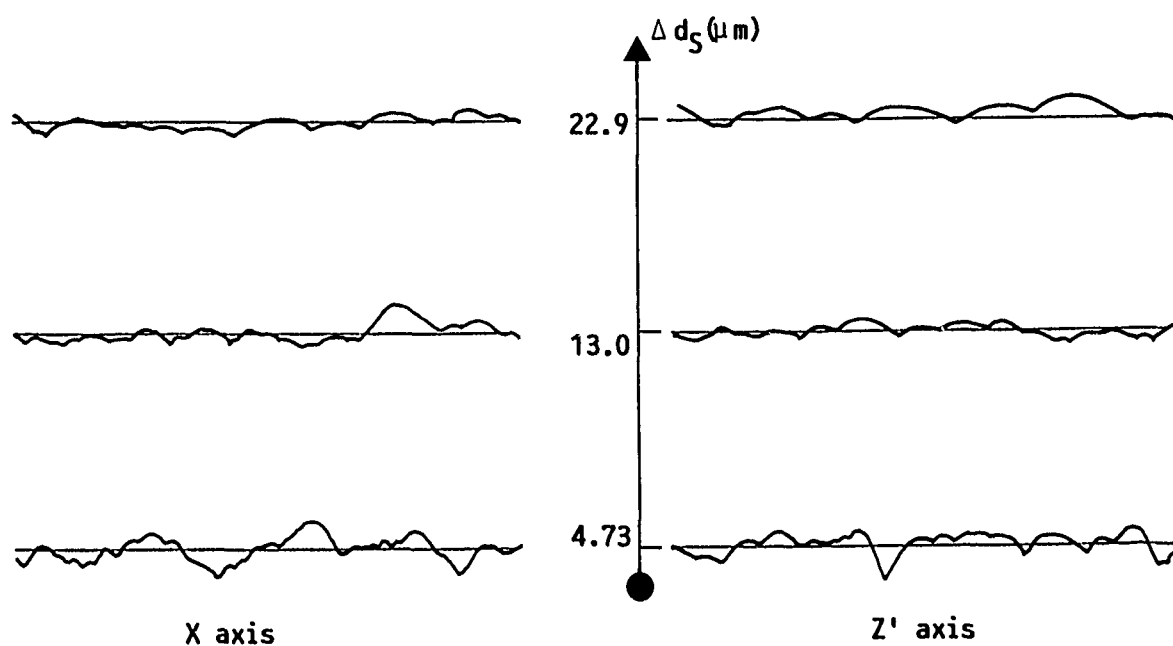


Figure 10 : Changes in the surface profilometry traces with the depth, Δd_S , of etch of an AT-49 quartz plate.

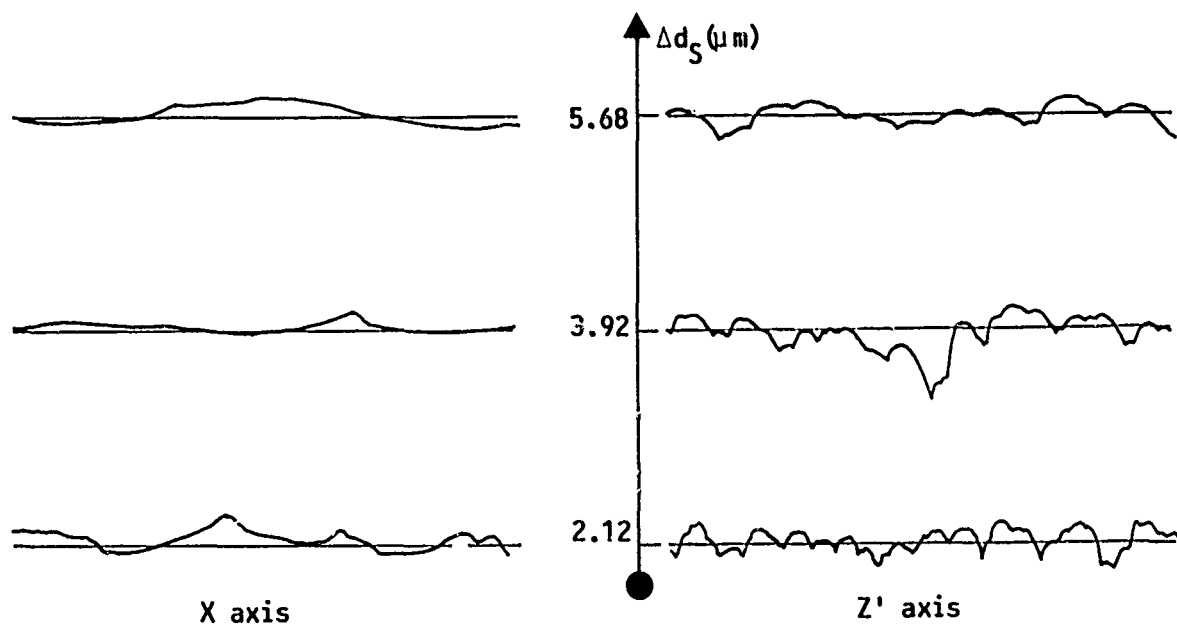


Figure 11 : Changes in the surface profilometry traces with the depth, Δd_S , of etch of an BT-19 quartz plate.

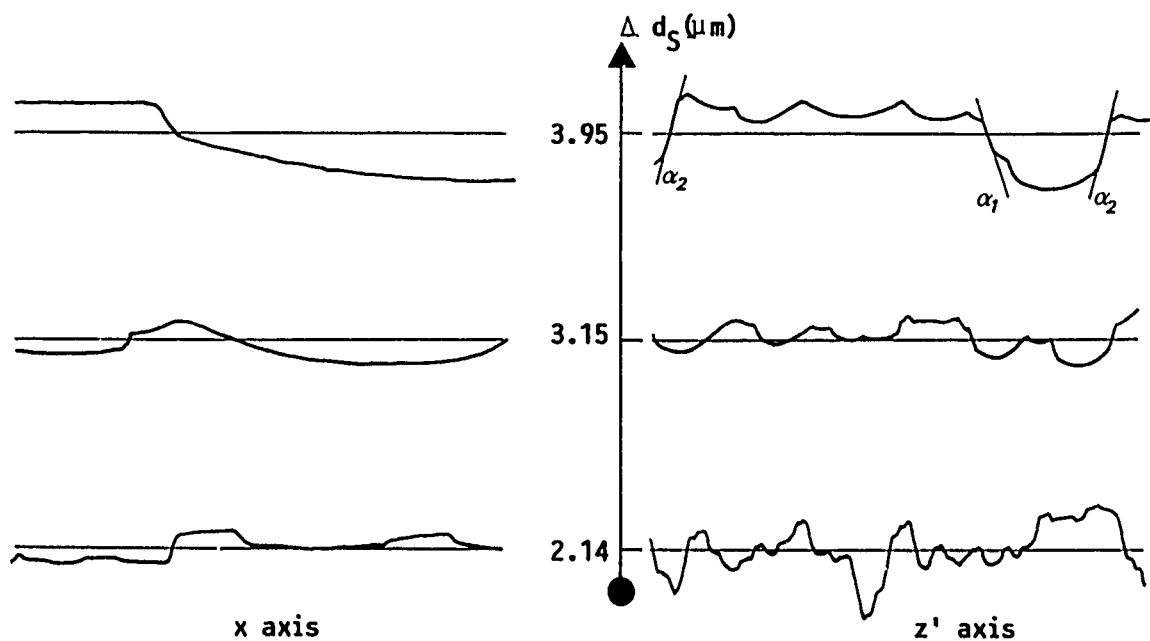


Figure 12 : Changes in the surface profilometry traces with the depth, Δd_s , of etch of an BT-39 quartz plate.

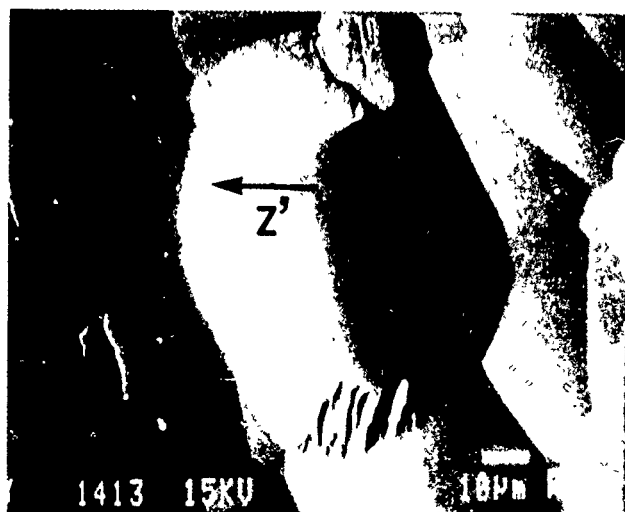


Figure 13 : Final SEM micrograph of a deeply etched AT-37 quartz plate.



Figure 14 : Final SEM micrograph of a deeply etched AT-49 quartz plate.

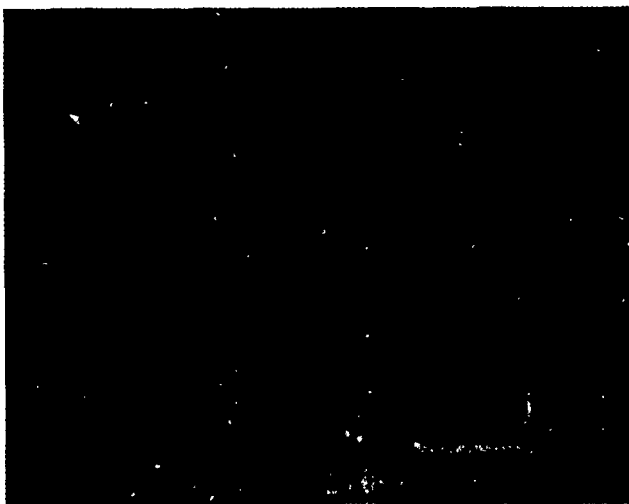


Figure 15 : Final SEM micrograph of an BT-19 quartz plate.



Figure 16 : Final SEM micrograph of an BT-39 quartz plate.

All these features are evidenced by the SEM micrographs. Effectively, figures 13 to 16 show that

1. The AT-37 surface is covered with uniformly shaped and slightly asymmetrical pits which are slightly elongated along a direction close to the X axis.
2. On the AT-49 surface develops convex terraces which extend in a direction close to the Z' axis.
3. Etching of the BT-19 surface reveals convex hillocks all elongated along the X direction.
4. Repeated etchings on the BT-39 surface give still different figures. Relatively flat bottomed and markedly elongated pits, all uniformly X oriented are formed. But individual pits are clearly bounded by sharp planes which extend in the X direction.

Discussion

Without attempting to undertake a detailed comparison of the predictions of the numerical simulation with the various features of differently oriented quartz surfaces revealed by repeated etchings^{5,10,13-15,18,19} a discussion of the typical results on etched BT-19, BT-39, AT-37 and AT-49 surfaces can constitute a convenient approach to establish the consistency of this mathematical model of etching.

Comparison of the changes in profilometry traces of the AT-37 surface on prolonged etching (Figure 9) with the etched profiles derived from the numerical simulation (Figure 6) shows a qualitative agreement : a concave background structure seems to result from the existence of a concavity around the reference surface in the polar diagram of the etch rate. The theoretical profiles of figure 7 agree well with the Z' profilometry traces of BT-19 and AT-49 quartz plates (Figures 10 and 11) : in all cases repeated etchings give rise to a convex background structure when the reference plane etches more rapidly than the adjacent planes. Moreover concave surface profiles with planar edges α_1 and α_2 have the tendency to develop when the orientation dependence of the etch rate is found to have the more complicated form shown in Figure 8. Thus we can establish a satisfactory correspondence between the shape of theoretical and experimental etch profiles as revealed respectively by Figures 8 and 12.

Thus various shapes of surface profile resulting from prolonged etching of differently oriented quartz crystals in ammonium bifluoride solutions are found to correspond to those obtained in the numerical simulation of moving surfaces derived from the Frank model. A relatively important property of the chemical attack of quartz crystals follows from this adequacy. Since the Frank dissolution theory is based on the hypothesis that the rate of dissolution of a crystal surface is a function only of orientation we infer that the etch rate of quartz crystal in the particular etchant considered is very closely completely orientation dependent. Hence this model provides satisfactory agreement with the qualitative description of the chemical attack mechanism and of the diffusion process which predict that the limiting concentration C_0 depends on orientation.

In the present application the numerical simulation was used to predict the shape of moving quartz surfaces selecting known etch rate distributions. But the converse approach is also possible. Hence the usefulness of the numerical simulation is also to be a tool to explore in details the variation of etch rate with orientation. Effectively the Z' profiles of singly rotated quartz plates present profile elements which are located in the YZ plane. As soon as the dependence of etch rate with the angle θ of rotation is known we can establish the relationship which describes the etch rate variation in various directions from a given reference surface. The shape of the etch figures in plan view will be then determined from this etch rate distribution. However to have valuable informations on the etch rate distribution it becomes necessary, as shown by various experimental works, to design resonators whose orientation varies very slowly. Hence an alternative approach to the etch rate distribution is

to compare the observed Z' etch profile on a quartz surface of given orientation with the results of the numerical simulation obtained for various etch rate-orientation relationships. This procedure seems to be more rapid than the long work which consists of designing a great number of quartz resonators.

Conclusion

Previous experiments^{9,10,13-15,18,19,33} on quartz plates of various orientations have revealed that the morphology of the final etch figures depends on the orientation. Thus a phenomenological model for the chemical attack is proposed in which two mechanisms are responsible to the existence of an orientation dependent etch rate, firstly the adsorption of a bifluoride ion at the quartz surface which involves that two O atoms expose simultaneously at the initial quartz surface and secondly the diffusion of the etchant toward the surface. According to the mechanism of adsorption a solution is presented for the continuity equation in which we consider that the etchant concentration tends to a limit characteristic to the crystal orientation.

Since these two mechanisms predict that the etch rate is governed by the crystal orientation analytical equations are derived in terms of the Frank model²⁹ in order to provide a two dimensional numerical simulation of the etched profiles of variously oriented quartz plates. The numerical simulation which distinguishes the diverging and converging trajectories gives evidence for particular etch rate relationships which result in the development of typical background structures.

The comparison of the theoretical moving profiles with the profilometry traces which characterize at various etching time the surface texture of some singly rotated quartz plates shows a complete agreement. Hence the Frank model provides satisfactory explanation and prediction for the dissolution of quartz in fluoride solutions. This agreement corroborates the fact that the dissolution mechanism is primarily determined by the crystal orientation. Such a conclusion is confirmed by previous experiments^{9,10,13,15,18,19}.

Moreover the numerical simulation also finds successful application to a more detailed evaluation of the variation of the etch rate, as a function of the orientation.

References

1. J.R. Vig, J.W. Lebus and R. Filler, Rep. ECOM - 4548, (U.S. Army Electronics Command, Fort Monmouth, NJ), (1977).
2. J.R. Vig, H. Wasshauser, C. Cook, M. Katz and E. Hafner, Proc. 27th Ann. Symp. on Frequency Control, Fort Monmouth, N.J., 1973, Electronic Industries Association, Washington, DC, p 98, (1973).
3. J.R. Vig, C.F. Cook, K. Schwidtal, J.W. Lebus and E. Hafner, Proc. 28th Ann. Symp. on Frequency Control, Fort Monmouth, N.J., 1974, Electronic Industries Association, Washington, DC, p 96, (1974).
4. H. Fukuyo, N. Oura, N. Kitajima and H. Kono, J. Appl. Phys., 50, 3653, (1979).
5. C.R. Tellier, Surf. Technol., 21, 83, (1984).
6. J.R. Vig, R.J. Brandmayr, R.L. Filler, Report No DELET-TR-80-5 (U.S. Army Electronic Research and Development Command, Fort Monmouth, N.J., 1980).
7. R.J. Brandmayr, J.R. Vig, Report N° DELET-TR-81-16 (U.S. Army Electronic Research and Development Command, Fort Monmouth, N.J., 1981).
8. A.J. Bernot, Proc. 39th Ann. Symp. on Frequency Control, Philadelphia, Pa, 1985, I.E.E.E., New York, N.Y., p 271, (1985).
9. R.J. Brandmayr, J.R. Vig, Proc. 39th Ann. Symp. on Frequency Control, Philadelphia, Pa, 1985, I.E.E.E., New York, N.Y., p 276, (1985).
10. C.R. Tellier, Proc. 39th Ann. Symp. on Frequency Control, Philadelphia, Pa, 1985, I.E.E.E., New York, N.Y., p 282, (1985).
11. C.R. Tellier and J.L. Vaterkowski, Surf. Technol., 26, 275, (1985).
12. C.R. Tellier, J. Mater Sci, 17, 1348, (1982).
13. C. Tellier, Proc. 38th Ann. Symp. on Frequency Control, Philadelphia, Pa, 1984, I.E.E.E., New York, N.Y., p 105, (1984).
14. C.R. Tellier, Proc. XIth Intern. Congress of Chronometry, Besançon, France, 1984, Société Française des Microtechniques et de Chronométrie, Besançon, p 115, (1984).
15. C.R. Tellier, F. Jouffroy and C. Buron, Mater. Chem. & Phys., 14, 25, (1986).
16. D. Ang, Proc. 32nd Ann. Symp. Frequency Control, Fort Monmouth, N.J., 1978, Electronic Industries Association, Washington, DC, p 282, (1978).
17. H. Fukuyo and N. Oura, Proc. 30th Ann. Symp. Frequency Control, Fort Monmouth, N.J., 1976, Electronic Industries Association, Washington, DC, p 254, (1976).
18. C.R. Tellier and C. Buron, Surf. Technol., 22, 287, (1984).
19. C.R. Tellier and F. Jouffroy, J. Mater. Sci., 18, 3621, (1983).
20. B.A. Irving in P.J. Holmes (Ed.), "The Electrochemistry of Semiconductors", Academic Press, London, pp 256-289, (1962).
21. F.M. Ernsberger, J. Phys. Chem. Solids, 13, 347, (1960).
22. H.N. Farrer and F.J.C. Rossotti, J. Inorg Nucl. Chem., 26, 1959, (1964).
23. J.S. Judge, J. Electrochem. Soc., 118, 1772, (1971).
24. V.V. Soroka, E.I. Lazorina and V.N. Stepanchuk, Sov. Phys. Crystallogr., 22, 353, (1977).

25. W.G. Palmer, J. Chem. Soc., 1656, (1930).
26. G.B. Larrabee, K.G. Heinen and S.A. Harrell, J. Electrochem. Soc., 114, 867, (1967).
27. Y. Adda and J. Philibert, "La Diffusion dans les Solides", Presses Universitaires de France, Paris, Vol. 1, Chapter 3, (1966).
28. H.K. Kuiken, Proc. R. Soc. London, A 392, 199, (1984).
29. F.C. Frank in R.H. Doremus, B.W. Roberts, D. Turnbull (Eds), "Growth and Perfection of Crystals", John Wiley, New York, pp 411-419, (1958).
30. N. Cabrera and D.A. Vermilyea in R.H. Doremus, B.W. Roberts, D. Turnbull (Eds.), "Growth and Perfection of Crystals", John Wiley, New York, pp 393 - 408, (1958).
31. F.C. Frank and M.B. Ives, J. Appl. Phys., 31, 1996, (1960).
32. R.B. Heimann in J. Grabmaier (Ed.), "Silicon Chemical Etching", Springer, Berlin, 197, (1982).
33. M.W. Wegner and J.M. Christie, Phys. Chem. Minerals, 9, 67, (1983).

* * * * *

This work is supported by the Direction des Recherches Etudes et Techniques which depends on the French Department of Defense.

* * * * *

FURTHER RESULTS ON THE USE OF SURFACTANTS IN CHEMICALLY POLISHING QUARTZ CRYSTALS

Ronald J. Brandmayr and John R. Vig

US Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, New Jersey 07703-5000

ABSTRACT

At the 39th Annual Symposium on Frequency Control, the authors reported on the use of surfactants in the etching solutions used to chemically polish SC-, AT-, and Z-cut crystals.¹ Surfactants could make the chemical polishing of AT- and SC-cuts "virtually foolproof." However, the Z-cut's chemical polishing remained critically dependent on the cleanliness of the surfaces prior to etching when the etching was performed at 75°C. Work on simplifying the chemical polishing procedures through the use of surfactants has continued and has been extended to AK-, ST-, and BT-cuts.

Various AK-cuts have been etched in hydrofluoric-acid-based solutions utilizing surfactants in the solutions. It has been found that most AK-cut crystals can be chemically polished in an $\text{NH}_4\text{F}:\text{HF} = 4:1$ solution. After investigation of the effect of surfactant additives on the contamination sensitivity of AK-cut crystals it has been determined that the contamination sensitivity of AK-cut crystals is very low compared with those of the SC- and Z-cuts. The ST- and BT-cuts also exhibit very low contamination sensitivity. Z-cut crystals were etched at temperatures ranging from 75°C to room temperature. The contamination sensitivity is less pronounced at 40°C than at 75°C. Therefore, it is significantly easier to chemically polish Z-cut crystals at 40°C than at 75°C. In no case did the use of surfactants eliminate the need for some precleaning of the Z-cut crystals.

INTRODUCTION

The objective of this paper is to report on the continuation of work presented at the 39th Annual Symposium on Frequency Control on the surfactant-assisted chemical polishing of quartz crystals.¹ Surfactant additives in the etching bath were shown to be effective in minimizing the contamination sensitivity of SC-cut and AT-cut quartz crystals during chemical polishing. Z-cut crystals, on the other hand, were nonresponsive to the use of surfactants under the conditions investigated. It was shown that concentrations in the range of 0.01 - 0.10 percent of the

proper surfactant were effective for the SC- and AT-cut crystals. In this paper, additional results on the Z-cut are reported. The evaluation of surfactants is also extended to the AK-, ST- and BT-cuts.

The variables of etching in surfactant-containing etching solutions include: etching temperature, surfactant type, surfactant concentration, etchant formulation, crystal precleaning procedure, and crystal angles of cut.

The qualities of etched surfaces were studied by means of optical inspection under a microscope at magnifications to about 30X, by scanning electron microscopy (SEM) and by profilometry. Surface profiles were measured using a Taylor-Hobson Talysurf. The surface roughness values were estimated by calculating the root-mean-square deviation from an imaginary center line through the profile, which was chosen so that the areas under the profile above and below the line were approximately equal. Alternatively, the roughness was estimated by digitizing with an HP plotter used in combination with an HP 9825A computer.

Table I shows the list of surfactants studied. An informative discussion on surfactants has been published by R. L. Camp et al.²

SURFACTANT	CLASSIFICATION ^{1,2}	CHEMICAL FAMILY ^{1,2}
FC-93, 95, 98, 99, 100	ANIONIC	FLUORO-CHEMICAL
ZONYL A	ANIONIC	HYDRO-CARBON
ZONYL FSN, FSD	NONIONIC	FLUORO-CHEMICAL
ZONYL TBS	ANIONIC	FLUORO-CHEMICAL
ZONYL FSC	CATIONIC	FLUORO-CHEMICAL
ZONYL FSK	AMPHOTERIC	FLUORO-CHEMICAL
LODYNE S-100	AMPHOTERIC	FLUORO-CHEMICAL
LODYNE S-103	ANIONIC	FLUORO-CHEMICAL
LODYNE S-107B	NONIONIC	FLUORO-CHEMICAL
SILVATOL NP	NONIONIC	HYDRO-CARBON
SUPERWET BOE 1235		PROPRIETARY

Table I - Surfactants Studied

Z-CUTS

Z-cut crystals have applications as seed crystals in growing cultured quartz, in sweeping, in resonators and in optics. The Z-cut can be chemically polished in $\text{NH}_4\text{F}:\text{HF} = 1:1$. However, because of its high contamination sensitivity, a thorough cleaning procedure must be employed before exposing the crystal to the etching solution; otherwise, smudging will develop and the crystal will not become polished. Attempts to employ surfactants in the etching solution to improve wettability and to obtain in situ cleaning, while successful for the SC- and the AT-cut, had not yielded positive results for the Z-cut when the etching was performed at 75°C. Contributing reasons for this may be the adsorption properties and the rapid etching rate of the Z-cut. If the etchant attacks the quartz faster than the surfactant attacks the surface contamination, then unevenly etched, smudged surfaces can result.

Table II shows a list of surfactants explored in order to obtain in situ cleaning of the Z-cut crystals and, thus, minimize the amount of precleaning required. Only a "quick water rinse,"

SURFACTANT ADDITIVE	CONCENTRATION	ETCHING TEMPERATURE	RESULTS AFTER "QUICK WATER RINSE"
FC-93	0.01 - 0.14 %	25, 30, 40, 45°	SMUDGED
FC-98	0.1%	75°	SMUDGED
FC-99	0.08 - 0.1 %	75°	SMUDGED
ZONYL A	1 ML	75°	SMUDGED
ZONYL FSN	0.1 %	35°	SMUDGED
ZONYL TBS	0.2 %	50°	SMUDGED
LODYNE 107B	0.1 %	35°	SMUDGED

Table II - Polishing results for Z-cut crystals.

i.e., a one-minute rinse under cold, running tap water was used to preclean the as-received crystals. This procedure was substituted for the more elaborate, but reliable, four-step process normally used. In no instance was a polished crystal obtained. Table III shows combinations of surfactants tried for the Z-cut. In the examples shown, a "quick-clean" was substituted for the standard four-step procedure. The "quick-clean" consisted of two steps: (1) ultrasonic cleaning in a cleaning solution³ for two minutes and (2) immersion of the crystals in boiling H_2O_2 for four minutes. This is less time-consuming than the four-step procedure normally used to clean the crystals. The "quick clean" resulted in polished crystals whenever the etching temperature was 45°C or lower, and resulted in smudged surfaces at 75°C.

SURFACTANT COMBINATION	CONCENTRATION	ETCHING TEMPERATURE	RESULTS AFTER A "QUICK CLEAN"
FC-99, FC-93	0.2, 0.02 %	40°, 45°	POLISHED
FC-99, FC-98, FC-95, FC-93	0.2 % EA.	35°	POLISHED
FC-99, FC-93	0.2 % EA.	25°	POLISHED
FC-99, FC-93	0.2, 0.01 %	40°	POLISHED
NONE		40°	POLISHED
NONE		75°	SMUDGED
FC-120, FC-98, FC-95	0.1, 0.1, 0.01 %	75°	SMUDGED
FC-98, FC-95	0.1 % EA.	75°	SMUDGED
NONE		58°	POLISHED
NONE		65°	POLISHED/SMUDGED

Table III - Polishing results for Z-cut crystals using surfactant combinations.

In an experiment to determine whether it was the surfactant or the lower etching temperature that was responsible for the improved polishing, it was found that a polished crystal could be consistently obtained at 40°C, even if no surfactant was used in the solution. While not causing any deleterious effect on the etching results, using the surfactant combinations shown does not appear to have the striking effect it does for the SC- and AT-cuts. In no instance was it possible to simplify the cleaning further than the two-step "quick clean," and still obtain a polished, smudge-free crystal. Thus, the contamination sensitivity of the Z-cut is less pronounced when the etching is performed at lower temperatures. The improved contamination sensitivity appears to be due more to the lower etching temperature employed than to the surfactants.

Additional etching was performed at temperatures between 40° and 75°C. It was found that a polished Z-cut could also be obtained at 58°C and 65°C without the use of surfactants, but not consistently. Some crystals became polished, others became smudged.

Figure 1 shows a comparison of two Z-cut crystals: one was polished at 40°C with FC-99 surfactant in the etching solution and the other without a surfactant at 40°C. No significant

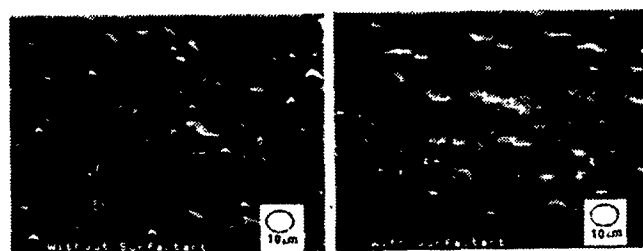


Figure 1. Two Z-cut crystal plates chemically polished at 40°C, with surfactant FC-99 and without surfactant.

difference was observed between the two surfaces. The Tallysurf profile for each crystal was about $0.16 \mu\text{m}$ surface roughness.

AK-CUTS

Kahan and Euler^{4,5} have reported results on doubly rotated quartz cuts with angles of cut in the ranges $30^\circ < \phi < 46.1^\circ$ and $21^\circ < \theta < 28.44^\circ$. These crystals are designated AK-cuts. The turnover temperatures for these cuts are relatively insensitive to crystallographic misalignments. AK-cuts having different angle combinations were obtained from Kahan and Euler and were etched for the purpose of developing solutions having the capability of chemically polishing these cuts. The cuts' contamination sensitivity, and the possible role of surfactants in the etching solution, were also investigated. The $\text{NH}_4\text{F}:\text{HF} = 4:1$ solution was selected as the initial candidate to polish the AK-cuts. Table IV shows a list of the angle combinations studied and the results obtained using this solution. Surfactant FC-99 fluorochemical was added to the etching solution, and the crystals were cleaned using the standard four-step cleaning process. All crystals were lapped $3 \mu\text{m}$ before etching. The etching temperature was 75°C . It can be seen that for the ten angle combinations shown, all became polished on both sides except the $(46.1^\circ, 23.58^\circ)$ crystal, which polished on one side only. A second attempt was made to polish this crystal by utilizing the Superwet BOE 10-1 solution; both sides of the crystal became polished after deep etching.

AK-CUT ANGLES (ϕ, θ)	ETCHANT WITH FC-99 ($\text{NH}_4\text{F}:\text{HF}$)	RESULT
$33^\circ/24.44^\circ$	4:1	POLISHED BOTH SIDES
$34^\circ/22^\circ$	4:1	POLISHED BOTH SIDES
$34^\circ/26^\circ$	4:1	POLISHED BOTH SIDES
$34^\circ/28.44^\circ$	4:1	POLISHED BOTH SIDES
$36.58^\circ/22^\circ$	4:1	POLISHED BOTH SIDES
$36.58^\circ/26^\circ$	4:1	POLISHED BOTH SIDES
$38.5^\circ/26^\circ$	4:1	POLISHED BOTH SIDES
$46.1^\circ/23.58^\circ$	4:1	POLISHED ONE SIDE
$30^\circ/24.45^\circ$	4:1	POLISHED BOTH SIDES
$36.58^\circ/28.45^\circ$	4:1	POLISHED BOTH SIDES
$46.1^\circ/23.58^\circ$	10:1*	POLISHED BOTH SIDES

*PROPRIETARY SURFACTANT.

Table IV - Polishing results for AK-cut crystals using FC-99 surfactant.

Figure 2 shows a photomicrograph of both sides of an AK-cut crystal having angles of cut of $(46.1^\circ, 23.58^\circ)$. The side on the right shows a poor surface roughness. When a similar crystal was etched in Superwet BOE 10-1, an improved result was obtained. Figure 3 shows both sides of an AK-cut having angles of cut $(38.5^\circ, 26^\circ)$. The difference in morphology is evident between the two sides of this crystal. Figure 4 shows two sides of an AK-cut having angles of cut of $(30^\circ, 24.45^\circ)$.

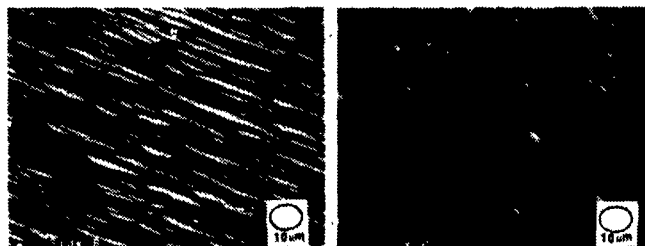


Figure 2. Two sides of an AK-cut deeply etched in $\text{NH}_4\text{F}:\text{HF}=4:1$ solution - $(46.1^\circ, 23.58^\circ)$.

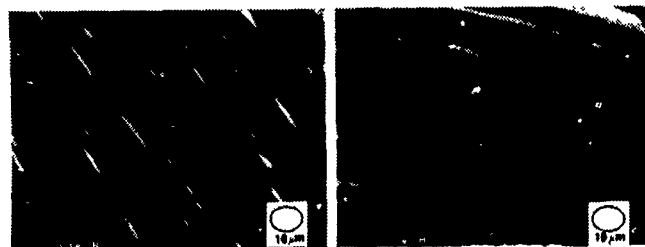


Figure 3. Two sides of an AK-cut deeply etched in $\text{NH}_4\text{F}:\text{HF} = 4:1$ solution - $(38.5^\circ, 26^\circ)$.

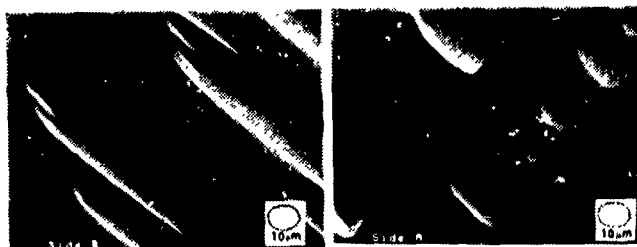


Figure 4. Two sides of an AK-cut deeply etched in $\text{NH}_4\text{F}:\text{HF} = 4:1$ solution - $(30^\circ, 24.45^\circ)$.

Crystals were then etched in $\text{NH}_4\text{F}:\text{HF} = 4:1$ solutions, this time without the use of surfactants and with only a "quick water rinse." The results are shown in Table V. A polished surface was obtained on all crystals, as before. The roughness values obtained with and without the surfactant are shown for both sides of the crystals. The good agreement in the surface roughness values indicates that a smooth surface can be obtained whether or not one uses a surfactant in the etching solution. The roughness values obtained are about that achieved for AT-cuts when etched in $\text{NH}_4\text{F}:\text{HF}$. Thus, the results

indicate that the AK-cuts are not contamination-sensitive when compared with SC- and Z-cuts. Only slight smudging could be observed when a "quick water rinse" was used to clean the crystals.

In 1979, chemical polishing results for ST- and BT-cut crystals were reported, but the contamination sensitivity of these cuts was not explored at that time.⁶ Therefore, it was, decided to reexamine the chemical polishing of these cuts.

POLISHING RESULTS FOR AK-CUT CRYSTALS
WITH AND WITHOUT SURFACTANT

AK-CUT ANGLES (θ, ϕ)	QUICK H ₂ O RINSE NO SURFACTANT	ROUGHNESS QUICK H ₂ O RINSE NO SURFACTANT	ROUGHNESS STD CLEAN FC-99
33°/24.44°	POLISHED BOTH SIDES	0.18 μ m, 0.25 μ m	0.16 μ m, 0.32 μ m
34°/22°	POLISHED BOTH SIDES	0.23 μ m, 0.15 μ m	0.13 μ m, 0.12 μ m
34°/28.44°	POLISHED BOTH SIDES	0.13 μ m, 0.25 μ m	0.07 μ m, 0.17 μ m
36.58°/22°	POLISHED BOTH SIDES	0.19 μ m, 0.18 μ m	0.091 μ m, 0.19 μ m
36.58°/26°	POLISHED BOTH SIDES	0.1 μ m, 0.2 μ m	0.019 μ m, 0.19 μ m
38.5°/26°	POLISHED BOTH SIDES	0.20 μ m, 0.18 μ m	0.27 μ m, 0.16 μ m
46.1°/23.58°	POLISHED BOTH SIDES	0.17 μ m, 0.46 μ m	0.19 μ m, 0.17 μ m
30°/24.45°	POLISHED BOTH SIDES	0.18 μ m, 0.30 μ m	0.22 μ m, 0.33 μ m

*SUPERWET BOE 10-1 ETCHANT USED FOR THIS CRYSTAL.

Table V - Polishing results for AK-cut crystals with and without surfactant.

ST-CUTS

ST-cuts have applications in surface acoustic wave devices. ST-cut crystals were obtained from two sources and were etched in solutions of $\text{NH}_4\text{F} \cdot \text{HF}$ and $\text{NH}_4\text{F}:\text{HF} = 4:1$. Table VI shows the results obtained with and without the use of surfactants in the etching solution. The ST-cuts were lapped 3 μ m before etching. The table shows that, when the ST-cut is etched with $\text{NH}_4\text{F} \cdot \text{HF}$ and FC-99 surfactant, a polished surface is obtained whether the standard cleaning procedure or, merely, a "quick water rinse" is used to clean the crystal. No smudging was observed after the "quick water rinse." An ST-cut was then intentionally contaminated with fingerprints and etched in $\text{NH}_4\text{F} \cdot \text{HF}$ without the use of a surfactant; a polished crystal also resulted. The ST-cut exhibited very little contamination-sensitivity. A polished ST-cut could also be obtained by using Superwet BOE 4-1 solution, but the etching rate was much slower than with the $\text{NH}_4\text{F} \cdot \text{HF}$.

Figure 5 shows a polished ST-cut crystal surface for which a surface roughness of 0.34 μ m was measured.

INTENTIONAL CONTAMINATION	ETCHANT AND SURFACTANT	QUICK WATER RINSE	STANDARD CLEANING
NO	$\text{NH}_4\text{F} \cdot \text{HF}$ FC-99; (0.1%)		POLISHED
NO	$\text{NH}_4\text{F} \cdot \text{HF}$ FC-99; (0.1%)	POLISHED	
YES	$\text{NH}_4\text{F} \cdot \text{HF}$ NONE	POLISHED	
NO	SUPERWET BOE 4-1 PROPRIETARY		POLISHED
NO	$\text{NH}_4\text{F} \cdot \text{HF}$ FC-99; (0.1%)		POLISHED

Table VI - Polishing results for ST-cut crystals.



Figure 5. ST-cut chemically polished with surfactant.

BT-CUTS

BT-cuts are used in bulk acoustic wave devices, and have been shown to have potential application in shallow bulk acoustic wave devices.⁷ The etchants used for the BT-cuts were $\text{NH}_4\text{F} \cdot \text{HF}$ and $\text{NH}_4\text{F}:\text{HF} = 4:1$. The crystals, which were made of natural quartz, were lapped 3 μ m before etching. Table VII shows the results of the experiments. It can be seen that the BT-cut can be polished, without smudging, in $\text{NH}_4\text{F} \cdot \text{HF}$ with FC-99 surfactant after only a "quick water rinse." When the BT-cut was intentionally contaminated with fingerprints and etched in $\text{NH}_4\text{F} \cdot \text{HF}$ without a surfactant, and after only a "quick water rinse" for precleaning, a polished surface was obtained, thus indicating a low contamination-sensitivity. Figure 6 shows two BT-cut surfaces. The first was etched in $\text{NH}_4\text{F} \cdot \text{HF}$ with surfactant FC-99 after a "quick water rinse" and the second was etched without a surfactant after the standard cleaning procedure. The surface roughness values were 0.08 μ m and 0.13 μ m, respectively.

INTENTIONAL CONTAMINATION	ETCHANT AND SURFACTANT	QUICK WATER RINSE	STANDARD CLEANING
NO	$\text{NH}_4\text{F} \cdot \text{HF}$ FC-99; (0.1%)		POLISHED
NO	$\text{NH}_4\text{F} \cdot \text{HF}$ FC-99; (0.1%)	POLISHED	
NO	$\frac{\text{NH}_4\text{F}}{\text{HF}} = \frac{4}{1}$ FC-99; (0.1%)	POLISHED	
YES	$\text{NH}_4\text{F} \cdot \text{HF}$	POLISHED	

Table VII - Polishing results for BT-cut crystals.

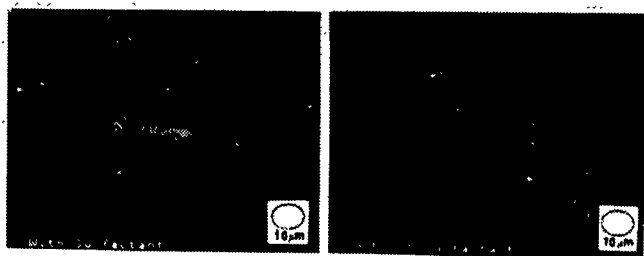


Figure 6. BT-cut chemically polished with surfactant FC-99 (left) and without surfactant (right).

SC-CUTS

Previously, it has been shown that, when chemically polishing SC-cuts, the use of FC-99 surfactant in the etching bath provided striking results in reducing the need for precleaning this contamination-sensitive cut. In an attempt to answer the question of whether the FC-99 surfactant leaves any residue on the etched crystal, four SC-cut crystals were etched for 30 minutes in a $\text{NH}_4\text{F}:\text{HF} = 4:1$ solution containing FC-99 surfactant. The crystals were then removed from the etching bath, rinsed, and spun dry in the usual manner. All four crystals were submitted for Auger and ESCA analyses, and were compared with a similar number of control samples. The results showed no differences in surface contaminants when the experimental samples were compared with the control samples, i.e., the surfactant left no measurable residue on the surface.

SUMMARY AND CONCLUSIONS

Z-cut crystals can be chemically polished at temperatures of 65°C and lower, and with a shortened "quick-clean" precleaning process, with or without surfactants in the etching bath. The contamination sensitivity of the Z-cut can be significantly reduced by etching at lower temperatures. Surfactants can be added to the $\text{NH}_4\text{F}:\text{HF} = 1:1$ solution without any deleterious effects. In no instance was it possible to chemically polish the Z-cut without some precleaning.

For the AK-cut, it was found that all but one angle combination investigated can be chemically polished in $\text{NH}_4\text{F}:\text{HF} = 4:1$. Superwet BOE 10-1 was needed in order to polish the (46.1°, 23.58°) crystal on both sides. The results of etching without the use of surfactants and with using only a "quick water rinse" demonstrate that the AK-cut has very low contamination-sensitivity.

ST-cuts and BT-cuts also have low contamination sensitivities. However, even for these cuts, the use of surfactants can provide a low cost insurance against uneven etching.

Auger and ESCA analyses results indicate that there is no residue from the FC-99 surfactant on etched SC-cut surfaces.

ACKNOWLEDGEMENTS

The authors thank the following people: Donald Eckart and Mary Lloyd-Saunders for photomicrographs, William Washington for the crystal lapping, Donald Fox for Auger and ESCA analyses, John Gualtieri for the HP 9825A program for estimating roughness, Ted Lukaszek for providing ST-cut samples, Thomas Parker, Raytheon Co., also for providing ST-cut samples and Joseph Balascio, Motorola, Inc., for providing Z-cut samples..

References

1. Ronald J. Brandmayr and John R. Vig, "Chemical Polishing in Etching Solutions that Contain Surfactants," Proc. 39th ASFC, pp. 276-281, 1985.
2. R. L. Camp, K. C. Scott, K. F. Scholne and R. R. Holland, "Success of Your Process May Depend on a Surfactant," Research and Development, March 1985, pp. 92-97.
3. Branson Ultrasonic Cleaner, Branson Cleaning Equipment Company, Parrot Drive, Shelton, Conn. 06484; Micro Liquid Laboratory Cleaner, International Products Corp., P. O. Box 118, Trenton, NJ.
4. F. Euler and A. Kahan, "Harmonic and Anharmonic Modes of AK-cut Crystal Resonators," Proc. 38th ASFC, pp. 150-156, 1984.
5. A. Kahan, "Turnover Temperatures for Doubly Rotated Quartz," Proc. 36th ASFC, pp. 170-173, 1982.
6. J. R. Vig, R. J. Brandmayr and R. L. Filler, "Etching Studies on Singly and Doubly Rotated Quartz Plates," Proc. 33rd ASFC, pp. 351-358, 1979.
7. A. Ballato and T. Lukaszek, "Shallow Bulk Acoustic Wave Program and Prospects," IEEE Trans. Microwave Theory and Techniques, MIT-27, pp. 1004-1012, Dec. 1979.

QUARTZ ANALOGUES

John H. Sherman, Jr.
2022 Woodcrest Drive
Lynchburg, VA 24503

Abstract

A family of 25 inorganic chemicals, five of which probably can never be prepared, is examined for the existence of alpha quartz analogues. The family results from oxidation of all of the binary III-V semiconductors. Record is found of the preparation of 12 members. Six were shown to be crystallographic analogues of alpha-quartz. The other six were shown to be analogues of higher temperature forms of silica. The paper speculates on the possibility of growing some of the latter six in the alpha-quartz form and on the prospects for the unexamined eight. Attention is given to hydrothermal growth in pressure vessels and to growth in non-aqueous fluxes.

Introduction

An interest in analogues of quartz arises from the conviction, or at least the expectation, that analogues of quartz will share with quartz some of the properties which have made of quartz such an important material. Quartz itself is such an unexpectable blessing to mankind that its very existence might well be recalled in any debate over the benevolence of the creator. From our own parochial viewpoint, let us remark only that its most interesting electrical properties exert themselves in the temperature range approximately between the freezing point and the boiling point of water, a range which is another of those unexpectable blessings of creation, the primary determiner of our own physical properties.

Quartz is the archetype. There is no other material of similar ubiquity with properties anywhere as attractive to the electronics community. Quartz does not do everything that we can imagine, however. After we have seen what quartz can do for us, we are sure we are needful of more. We have searched among esoteric piezoelectric materials we know how to make, and a majority of the crystal structures are capable of piezoelectricity, for other materials which will be just like quartz, only more so. What we seek contains the set of requirements that the material must demonstrate a good mechanical Q, display mechanical resonance modes having low frequency variation in a range of temperatures matching the operating temperature ranges of electronic devices, and that the electro-mechanical coupling be other than that of quartz, usually stated as greater than that of quartz.

Any material which looks like quartz will be at once attractive as a candidate material because we recognize it as an analogue of quartz, and we

naively expect it to behave in an analogous manner in those respects which make us value quartz.

Analogizing

Reasoning by analogy is surely the most primitive form of reasoning. It is surely also one of the most fruitful. In undertaking a conservative approach to solving a problem, the discovery of another problem already solved which can serve as a model is surely the most expeditious beginning. If the truth were known, reasoning by analogy is probably the first technique used by almost every successful investigator. To expect a material to act like quartz just because it looks like quartz is asking a lot, but we should plan to be lucky. If we turn out to be lucky, we have thereby saved a lot of time.

Having analogized that we want to look at things like quartz, we might well look first at things which have a simple family relationship to quartz. This is another kind of analogizing. If we want to find analogues for quartz, therefore, the first place to look would probably be at the sulfide, selenide and telluride of silicon and to the oxides of carbon, germanium, tin and lead. To make the search complete, all combinations should be considered. Thus we should consider the possibility that carbon dioxide might crystallize in the same system, and, failing that, carbon disulfide. After only a little looking at the 20 compounds so implied, only two are found really to look even a little like quartz. One is germanium dioxide, which crystallizes similarly and has been studied, but has not appeared attractive for other reasons. The other looks very much like quartz, and, in fact, is quartz. That is to say, silicon dioxide is one of the family, and it can, of course, be made to crystallize in the form of alpha quartz.

In a strange way the perception has forced its way into my consciousness that the first analogue of quartz investigated in any depth was the industrially crystallized silica we call by the name "cultured quartz". Quartz is, and has for all but an almost negligible fraction of the time of its use by man, been a natural material. Any material offered as a substitute for this must approximate, at least, the properties of quartz as found in nature. For the first twenty years the industrial product was an incompletely adequate substitute for real quartz, though it was a most promising analogue. At long last it is almost completely adequate as a substitute. No other

analogue can be expected to approach quartz as completely as this. On the other hand, other analogues may turn out to have properties sufficiently different from those of quartz that they have value in their own right, yet be sufficiently similar that our 100 years of experience with quartz may save us much valuable time in the exploitation of them.

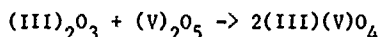
Analogues of Silicon

The examination of the periodic table for the elements which lie in chemical family relationship with silicon and oxygen has already been done seeking analogues, but without success. Probably in some important sense carbon and germanium are not adequately analogues of silicon, and sulfur is not an adequate analogue of oxygen. Germanium just barely misses. Its oxide comes quite close to being a useful analogue of quartz, but has already been examined and rejected.

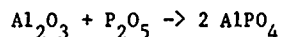
The question now arises just what an analogue of silicon should be like. Perhaps the key is not simply to lie in the same column in the periodic table. Both silicon and germanium are semiconductors and share the same cubic crystal structure. Perhaps the proper key to analogy is the semiconducting property. If this is a proper key for our analogizing, then there exists another family of materials created for no other reason than to act as analogues of semiconducting silicon. They have been coerced into crystallizing into the cubic form of silicon and are semiconductors. If they are the proper analogues of silicon then their oxides might make analogues of quartz.

These materials are the III-V semiconductors. They are compounds of elements which are themselves members of families which have a column relationship in the periodic table, which kind of relationship failed to generate a rich find of quartz analogues in the earlier try. They have been the subject of intense study in the semiconductor industry. They have been most interesting analogues of silicon, in that they variously display properties inherent in the silicon kind of crystal, but with those properties emphasized quite differently than we see in silicon itself. As analogues of silicon they display something like the range of properties which we might hope to occur in a family of analogues of quartz.

As quartz is silicon completely oxidized, the analogues would be the III-V materials completely oxidized. In simple chemical nomenclature we would write an expression of the kind:



as a description of the material synthesis we are looking for. To be explicit, in specifying the kinds of compound defined, we should be able to order up, in Chinese restaurant style, one from column III and one from column V, thus to get one of our candidate compounds. For instance for the III element we might select aluminum and for the V we might select phosphorus. In this case we would be considering the compound:



The III and V Elements

The columns of the periodic table are a code for the quantum configuration of the electrons in the outside or non-completed shells of electrons of the various elements. In particular, the elements in any column are characterized by their valence electrons. In the lighter elements there is no ambiguity in the family relationships. The elements simply walk up a ladder through two cycles of eight elements as two outer shells of eight electrons, the K and L shells are filled. The heavier elements display sub cycles as shells M, N, O, P and Q are filled. The ones we are tracking are the ones in the III-A column having two s and one p electrons and the ones in the V-A column having two s and three p electrons, respectively. These are the same elements which are combined to make up the significant semiconductor combinations.

The list of III-A elements comprises: Boron, Aluminum, Gallium, Indium, Thallium. The list of the V-A elements comprises: Nitrogen, Phosphorus, Arsenic, Antimony and Bismuth. With five of each kind, there are 25 binaries and 25 complete oxides of those binaries. They are tabulated here:

BNO ₄	BPO ₄	BAso ₄	BSbo ₄	BBio ₄
AlNO ₄	AlPO ₄	AlAsO ₄	AlSbo ₄	AlBio ₄
GaNO ₄	GaPO ₄	GaAsO ₄	GaSbo ₄	GaBio ₄
InNO ₄	InPO ₄	InAsO ₄	InSbo ₄	InBio ₄
TlNO ₄	TlPO ₄	TlAsO ₄	TlSbo ₄	TlBio ₄

Table I, The III-V Oxides

Impossible or Improbable Compounds

The left column of the table of III-V oxides looks queer. The valences are right but nitrates just do not form this way. Although fully oxidized Nitrogen has the formula N₂O₅, the acid is HNO₃, not H₃NO₄. Nitric acid is spontaneously dehydrated in contrast with the other acids of its family. If the compounds (III)NO₄ could be made to form somehow, they could be expected not to be stable, at least in processing which would involve water. Thus they might not be expected to crystallize out of water solution, as do quartz and berlinite. Direct reaction of a III oxide with N₂O₅ might be possible, but it would have to be a slow reaction, due to the instability of N₂O₅ at elevated temperatures. At the very best, a means of synthesizing any compound of the first column is not immediately evident.

The International Centre For Diffraction Data

To facilitate research in unfamiliar materials there is maintained a central registry of the X-ray diffraction spectra of all kinds of solids. This is maintained at the International Centre for Diffraction Data [1]. It was in the data base maintained by this establishment that the table of III-V oxide compounds was checked out.

As was anticipated, they have no record of the strange nitrates having ever been prepared. At least no data on the powder pattern diffraction spectrum for any of those compounds was to be found. Of the others, there are no entries for any of the bismuth compounds, nor for BSbO_4 , TlAsO_4 , or TlSbO_4 . There are, therefore, records for 12 of the compounds. For several there are reports of the results of several methods of synthesis and other processing.

Six Demonstrated Quartz Analogues

Six of the twelve are reported to have structures analogous to alpha quartz. They are the phosphates and the arsenates of boron, aluminum, and gallium. Aluminum phosphate is familiar to us. Much effort has been expended on the study of this material since the early 1950s. The others are, I believe, strangers to us.

Boron phosphate and arsenate are both found crystallized in the tetragonal system and in a hexagonal form called "isostructural with quartz". At least one other high pressure hexagonal form of each exists, which can be prepared in a hydrothermal bomb at 85000 atmospheres and a temperature of 1200 degrees C.

Aluminum phosphate is reported in a variety of forms almost as diverse as silica. There are several tridymite forms, two other orthorhombics, high and low cristobalites and analogues of alpha- and beta- quartz.

The original substantial study of aluminum phosphate in its alpha-quartz configuration was a project conducted at the Signal Corps Engineering Laboratories by J. M. Stanley, reported in 1954 [2]. Stanley's article does not state how aluminum phosphate was chosen for the project. My impression is rather strong that the ultimate reason so much effort has been expended on Berlinite is not any peculiar merit of the material itself, but that Stanley's work was brilliant and successful, and that the large crystals he grew were made available for distribution and study. No one made a similar effort on any of the others.

Berlinite is one of the quartz analogues named by Egli [3] in his Survey of Inorganic Piezoelectric Materials. Egli listed 16 materials of crystal structure supposedly analogues of alpha quartz. Three of them have water of crystallization. The formulas of most of the rest are at least surprising, of this more later. He passed up some of the proven members of our family as not analogues and as not piezoelectric. If Egli's paper was the source, the selection of aluminum phosphate from it was most fortuitous.

Aluminum arsenate is reported in tetragonal and in alpha-quartz forms.

Aluminum phosphate is recorded to show an alpha-beta phase transition at 579 degrees C the equivalent of that of quartz at 573 degrees. Other information [4] denies the existence of this transformation in gallium phosphate below its

temperature of conversion to a cristobalite structure above 935 degrees, and below 900 degrees C in the arsenates of aluminum and gallium.

Someone, I cannot remember who, once said somewhere within my hearing, but just where or when I cannot recall, that there is a necessary connection between this phase transition and the possibility of positive temperature coefficients of frequency, so that if there is no alpha-beta transition there is no AT resonator. Ballato [5] does not know such a relation and thinks that the report is some kind of old wives tale. I have great respect for old wives. My grandmother was an old wife and she knew to put moldy bread poultices on my festering skinned knees in the early 1920s, a full generation before Fleming. In any case, here are three quartz analogues for study, in which it is reported that the transition does not occur. If AT- behavior occurs in any of them the rumor can be laid to rest.

Gallium phosphate is reported in analogues of both high- and low- cristobalite, as well as in an analogue of alpha-quartz. Gallium arsenate is reported only in the alpha-quartz form.

All six of these compounds are described as made or processed in hydrothermal pressure vessels.

The Six Non-Analogues

Data on the phosphates of indium and thallium are quite similar. The NBS prepared the samples hydrothermally at 350-400 degrees C and reported very rich powder patterns indicating orthorhombic structures. There is no reason to expect that alpha-quartz analogues can be made, at least hydrothermally, the conditions generating these samples being so similar to those which would be necessary to grow the analogues if they were to exist.

Indium arsenate was not adequately observed. The entry consists of a reference from a Russian journal containing a number of lines of powder pattern without identification of the crystal class. No information was given on the method of synthesis. This compound should be examined further.

The antimonates of aluminum, gallium and indium are reported to be of tetragonal or monoclinic habit and are compared to rutile. Nothing is given on the methods of synthesis or of any other processing. It cannot yet be concluded, therefore, that there are no antimonate analogues.

Of these six compounds indium arsenate and the antimonates deserve more study. Table II will show that along with one untested possibility they must not yet be completely rejected.

The Untested Compounds

In order to see the possibilities remaining in the remaining eight compounds, the matrix of Table I is expanded to contain a review of the

aforestated results for the twelve examined. This constitutes Table II. The nonexistent nitrates are omitted from Table II.

The table displays a cell of six compounds which crystallize as analogues of quartz. These are flanked by compounds which have been synthesized and examined, and which are not analogues. Because the antimonates were made by methods unstated, some possibility may remain among them. The bismuth and thallium compounds constituting the outer border of the matrix cannot be expected to provide analogues since they are separated from the analogues by a row and a column in which analogues do not appear.

Boron antimonate should be examined first. If it can be crystallized in analogue form then aluminum antimonate might also. If Boron antimonate is an analogue, then indium arsenate probably is not, but it should not be ignored, as it lies on the border at the corner, and we do not know the shape the corner should have.

Observations and Speculations

The semiconducting binary compounds are not simply mixtures of the two components in equal atomic concentrations. To be useful semiconductor materials they must be crystallized in very exact form with whole planes one atom thick of a single constituent alternating with whole planes one atom thick of the other constituent. Persuading them so to do is one of the reasons crystals of these materials are so difficult to grow. It is quite possible that the same kind of absolute alternation is also necessary in the crystals of the oxides. The unit cell of these quartz analogues consists of two almost equal "half cells", each derived from half an oxide molecule. In the crystal of Berlinite it is thus not completely meaningless to try to associate particular oxygen atoms with aluminum or with phosphorus. In quartz it is truly meaningless so to associate. This may contribute to the difficulty of growing crystals of Berlinite and may also stand in the way of growing any of the others.

Since most of these materials have been shown to be susceptible to hydrothermal synthesis and processing, they will probably be so treated, but this is not clearly necessary. There may be superior fluxes for the growing not yet sought out. There are materials which may make excellent fluxes which have much lower vapor pressures at the appropriate temperatures. Among these are RbOH (melting point 300 degrees C) [6], eutectics of PbO/PbF_2 (490 deg. C), NaF/NaPO_3 (490 deg. C), $\text{LiPO}_3/\text{KPO}_3$ (518 deg C) and $\text{NaPO}_3/\text{KPO}_3$ (535 Deg C) [4]. A major attraction of this kind of flux is that the process would take place at approximately atmospheric pressure rather than at the extreme pressure of the hydrothermal autoclave.

In this connection it must be pointed out that there is a significant error in [6] which must be corrected. This document would lead you to believe that excellent cultured quartz has been grown already in fused RbOH. This is, however an

error [7]. The crystals were grown in 0.5 molar RbOH in water in an otherwise normal hydrothermal process. This was one of a set of experiments which also included using similarly dilute solutions of CsOH and KOH [8]. The misapprehension might well be pursued to see if fused RbOH will prove a desirable flux after all.

There are surely other trivalent metals which can be tried as the III element in a III-V combination. Among these are ferric iron, chromic chromium, ceric cerium and manganic manganese. The phosphates of these metals all have water of crystallization, and might for this reason be mistrusted. This is not properly a disqualification. Almost always water of crystallization has evaporated by 200 deg. C. Available forms of aluminum phosphate are hydrated and the derived dehydrated forms are very hygroscopic, as are most forms of silica and silica gel. Ferric orthophosphate and arsenate might well prove to be useful quartz analogues, for example, as might the ceric, chromic or manganic salts.

Other possibilities are mentioned by Frondel [9]. He wants to consider anything of the formula AX_2 as a possible analogue, offering BeF_2 as an example. He describes all of his candidates as "weakened" models, meaning that critical temperatures tend to be lower for the analogues than for quartz. That this is a problematic generalization is displayed by the fact that the alpha-beta transition temperature of Berlinite is higher than that of quartz. Frondel mentions some, but not all of the compounds in this generating family.

Earlier on I remarked that the formulas of some of Egli's compounds proposed to be potential analogues of quartz were surprising. The problem I encountered in accepting them is that the formulas were of the form AX_3 instead of AX_2 . The justification of AX_2 is rather straightforward, considering SiO_2 , but AX_3 seems simply wrong. It may be correct but it is nevertheless surprising.

An industry exploiting the arsenates would certainly have to develop special techniques to evade the toxic nature of arsenic. We have a successful beryllium industry and a huge nuclear materials industry in which problems of handling safety of even greater magnitude have been faced and solved. If the properties of the arsenate quartz analogues prove advantageous, the industry will learn to handle them.

Conclusion

The generating concept, that the oxides of the cubic III-V semiconductor materials might be analogues of quartz, has led to the recognition of several analogues. None were found which were not known somewhere, but the principle of the search helped to organize a group of known materials into a related group. From this point further analogizing leads to the recognition of yet other candidate materials similarly related.

References

- [1] International Centre for Diffraction Data, 1601 Park Lane, Swarthmore, PA 19081
- [2] J. M. Stanley, "Hydrothermal Synthesis of Large Aluminum Phosphate Crystals", Industrial and Engineering Chemistry, Vol 46 no 8, Aug, 1954, pp 1684-1689
- [3] P. H. Egli, "A Survey of Inorganic Piezo-electric Materials", American Mineralogist, 33: 622-633 (1948)
- [4] G. A. Slack private communication, 1977
- [5] A. Ballato private communication, 1985
- [6] Union Carbide Corporation, Nuclear Division P. O. Box X, Oak Ridge Tn 37830, Industrial Cooperation Bulletin NTN-77/0297, 1977
- [7] O. C. Kopp private communication, 1986
- [8] O. C. Kopp and G. W. Clark, "Hydrothermal Synthesis Optical Perfection and Surface Topography of Quartz Grown in RbOH and Other Alkali Hydroxides", Journal of Crystal Growth 2 (1968) 308-312
- [9] C. Frondel, "Dana's System of Mineralogy", 7th edition, Vol III, 1962, Wiley, pp 5-7

Table II. The III-V Oxides Stable at Room Temperature as Filed With the International Centre

	Phosphate	Arsenate	Antimonate	Bismuthate
Boron	BPO ₄ (1) System:Tetragonal Prep:Not given (2) System:Hexagonal "Q structure type" Prep:Not given (3) System:Hexagonal "Isostructural w/Q" Prep:"High pressure" (4) System:Hexagonal Prep:1000-1200 deg C 85000 Atm	BAso ₄ (1) System:Tetragonal Prep: Not given (2) System:Hexagonal "Isostructural w/Q" Prep:"High Pressure" (3) System:Hexagonal Prep:1000-1200 deg C 85000 Atm.	BSbo ₄ (1) No report	BBio ₄ (1) No report
				Note: Q = Quartz HT = Hydrothermal
Aluminum	AlPO ₄ (1) System:"Tridymite" Two forms Prep:HT, 950 C (2) System:Orthorhombic Prep:Al, H ₃ PO ₄ , HT, 1200 C (3) System:Hexagonal "Q Group" Prep:Al, H ₃ PO ₄ , HT (4) 5 forms of undeter- mined system formed by dehydrating varis- cite.	AlAsO ₄ (1) System:Tetragonal "Rutile structure" Prep:HT, 900 C (2) System:Hexagonal "Low Q Structure" Prep:Al ₂ O ₃ , As ₂ O ₅ HT, 260°C, 50 Kbar	AlSbo ₄ (1) System:Tetragonal "Rutile structure" Prep:Not given	AlBio ₄ (1) No report
Gallium	GaPO ₄ (1) System:Orthorhombic "Low Cristobalite" Prep:Ga ₂ O ₃ , H ₃ PO ₄ , 1300 C (2) System:Hexagonal "Alpha Q structure" Prep:Ga, HPO ₃ , HT	GaAsO ₄ (1) System:Hexagonal "Low Q structure" Prep:Ga ₂ O ₃ , As ₂ O ₅ , HT, 260 C, 50 bar	GaSbo ₄ (1) Tetragonal "Rutile-like cell" Prep:Not stated	GaBio ₄ (1) No report
Indium	InPO ₄ (1) System:Orthorhombic Prep:In ₂ O ₃ , H ₃ PO ₄ , HT, 350-400 C	InAsO ₄ (1) System:Undetermined Prep:Not stated	InSbo ₄ (1) System:Monoclinic "Rutile-like cell" Prep:Not stated (2) System:Monoclinic "High pressure form" Prep:Not stated	InBio ₄ (1) No report
Thallium	TlPO ₄ (1) System:Orthorhombic Prep:Not Stated	TlAsO ₄ (1) No report	TlSbo ₄ (1) No report	TlBio ₄ (1) No report

STUDY OF IRRADIATION EFFECTS IN QUARTZ CRYSTALS USING LOW-TEMPERATURE DIELECTRIC RELAXATION

S. Ling and A. S. Nowick
Henry Krumb School of Mines, Columbia University
New York, NY 10027.

Summary

A study is made of irradiation effects on α -quartz crystals using the technique of dielectric loss measurements at low temperatures. Sodium-swept quartz shows a pair of loss peaks (at 30 K and 75 K for a 1 kHz frequency) that are due to the Al-Na defect. These peaks generally decrease after irradiation and a new "irradiation peak" appears in the range of 8 to 12 K. The fact that this "irradiation peak" also appears in a vacuum swept sample and in the same proportion as Al-hole centers strongly shows that this peak is related to the Al-hole center. Measurements of the two Al-Na peaks and the irradiation peak permits us to follow defect changes in quartz as a result of X-ray irradiation and subsequent annealing. Restoration of the main Al-Na peak during annealing occurs in two stages: one near 500 K and the other above 600 K. A defect model interpreting these two stages is presented.

Introduction

The relation of lattice defects to frequency instabilities of quartz crystal resonators, both before and after irradiation, has been much discussed in these Symposia[1] and elsewhere.[2,3] The purpose of the present paper is to show that dielectric loss measurements at cryogenic temperatures constitutes an important tool for the study of defects in quartz. Especially when used in combination with other techniques, such as infrared (IR) absorption, electron spin resonance (ESR) and anelastic relaxation, it is able to give considerable information about the defect structure of quartz crystals.

The principal defects present in as-grown cultured quartz crystals center about Al^{3+} impurities substituting for Si^{4+} . Because of the charge difference, Al^{3+} ions must be compensated by monovalent ions, e.g. Li^+ , Na^+ or H^+ , located in interstitial sites adjacent to the substitutional Al. The Al-Na pair manifests itself strikingly through two dielectric loss peaks at 30 K and 75 K (for a frequency of 1 kHz) known as the α and β peaks, respectively.[4,5] This defect also produces an analogous pair of mechanical loss, or anelastic relaxation, peaks.[6,7] It is known that the α and β defects involve the Na located on opposite sides of the distorted AlO_4 tetrahedron.[5,8] Since the Na interstitial site is located off the two-fold (C_2) symmetry axis (the x-axis), there are two equivalent α sites and two equivalent β sites on any one tetrahedron, as shown in Fig. 1. The dielectric and anelastic relaxation processes are due to the

reorientation of the Na^+ between either pair of equivalent sites. Since the α site is more highly occupied than the β , it is reasonable to expect that there is a free energy difference, Δg (>0) such that, under equilibrium conditions at temperature T , the

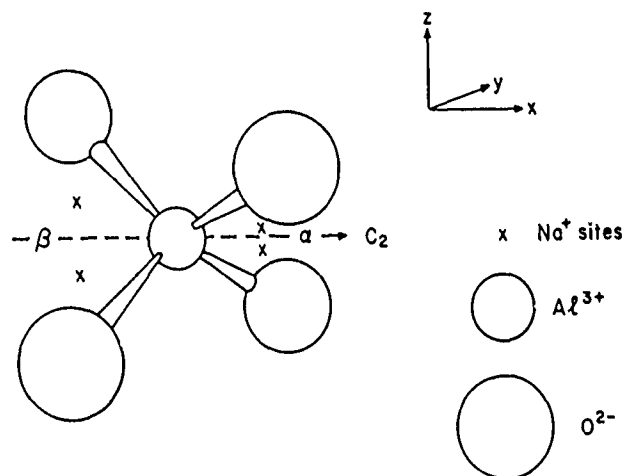


Fig. 1. Schematic diagram showing the distorted tetrahedron, the basic structural unit of α -quartz, with Al^{3+} replacing Si^{4+} . The tetrahedron contains a single twofold symmetry axis (in the x direction) designated C_2 . Also shown are the two equivalent α -sites and the two equivalent β -sites. An interstitial Na compensating the Al^{3+} will go into one of these four sites.

occupation ratio will be given by:

$$N_{\alpha}/N_{\beta} = \exp(\Delta g/kT) \quad (1)$$

In the case of Li^+ compensation, an Al-Li pair is formed, but in this case the Li sits on the C_2 axis; accordingly no dielectric or anelastic relaxation is observed.[7,8] Since Li is present in the mineralizer during growth of most cultured crystals, the Al-Li defect tends to predominate. However Li can be exchanged for Na by electrodiffusion or "sweeping" with a Na salt at the anode.[6,7] In such a Na-swept sample most Al is in the form of Al-Na pairs. The alkalis can also be replaced by H^+ through sweeping in moist air, to produce Al-OH centers.

The remaining important defects in quartz are the "grown-in OH centers" which manifest themselves through IR absorption bands near 3500 cm^{-1} . [9-11] These centers are different from Al-OH, but the nature of these defect centers have not yet been established.

Irradiation of quartz crystals with high-energy photons (X-rays or γ -rays) produces large numbers of electron-hole pairs. When irradiation is carried out above 200 K, it gives rise to changes in the defects, such that alkalis originally paired with Al are replaced by H^+ (as Al-OH centers) or holes (to form Al-h centers). (The former are detected by IR and the

latter by ESR.[12,13]) In addition, upon irradiation a dielectric relaxation peak is also observed near 10 K. Originally this peak was attributed to Al-h centers[14] but later results suggested that perhaps it might be due to an alkali center.[15] To keep open its origin, we simply refer to it as the "irradiation peak". In the present work we follow the production of this peak by irradiation and its disappearance during annealing. At the same time, by using Na-swept samples, we can follow the concurrent changes in the α and β peaks. In this way, we hope to learn more about defect processes that take place during irradiation and annealing.

In addition, since Al-h centers have been found in vacuum-swept samples without irradiation[16], we wished to examine such samples to see if the "irradiation peak" is also present under these conditions. In this way, we have an opportunity to establish the identity of the irradiation peak.

Theory

A defect that has lower point symmetry than its host crystal has several crystallographically equivalent orientations, among which it can reorient preferentially in the presence of an electric field.[17] If a sinusoidal electric field with angular frequency ω is applied to the system, the reorientation of the defect among its equivalent orientations give can give rise to one or more peaks in dielectric loss, $\tan \delta$, which obey the Debye equation:

$$\tan \delta = \epsilon''/\epsilon' = (\delta\epsilon/\epsilon_\infty) \cdot (\omega\tau / 1 + \omega^2\tau^2) \quad (2)$$

where ϵ'' , ϵ' are the imaginary and real parts of the complex dielectric constant, ϵ_∞ the high frequency dielectric constant, $\delta\epsilon$ the relaxation of the dielectric constant, and τ the relaxation time, often given by:

$$\tau^{-1} = \nu_0 \exp(-E/kT) \quad (3)$$

Here E is the activation enthalpy, kT its usual meaning, and ν_0 the pre-exponential factor. For a field parallel to the z -direction (the c -axis), the maximum peak height is given by:

$$\tan \delta_{\max} = \delta\epsilon / 2\epsilon_\infty = N_d \mu_z^2 / 2\epsilon_0 kT \quad (4)$$

where N_d is the concentration of the defect (in number/volume), μ_z is the component of the defect dipole moment in the z -direction, and ϵ_0 is the permittivity of free space.

For the Al-Na defect in quartz crystal, as already mentioned, two Debye peaks denoted as the α and β peaks are observed. A useful and measurable quantity, the ratio of the α to β peak heights, can be deduced from equation (4) to be:

$$R = \tan \delta_{\max}^\alpha / \tan \delta_{\max}^\beta = (N_\alpha / N_\beta) \cdot (T_\beta / T_\alpha) \cdot (\mu_3^\alpha / \mu_3^\beta)^2 \quad (5)$$

where T_α and T_β are the two peak temperatures, and μ_3^α and μ_3^β are the respective z -component of the dipole moments. Previous work showed that $T_\beta / T_\alpha = 75/30$ while $\mu_3^\alpha / \mu_3^\beta = 0.5(\pm 0.1)$. [8] Thus, from the ratio R , we can obtain N_α / N_β , but in absolute accuracy only to within $\pm 40\%$.

Table I. Crystals used in this work.

Designation	Source	Al content (ppma)
PQE	Sawyer Co. (U.S.)	17
QA26	U.S. Air Force RADC	55
H29-14	U.S. Air Force RADC	4
NQ	Natural (U.S.)	53

Experimental Details

Crystals Used

Various cultured and a natural crystals were used in this study. The different crystals used, with their designations and sources, are listed in Table I. The Al contents, given in the last column, were obtained from the heights of the Al¹⁴Na dielectric loss peak heights[8], except that for the vacuum swept H29-14 which was obtained from ESR measurement made by Halliburton at Oklahoma State University.[16] Na -sweeping of the crystals were carried out at Oklahoma State by Martin, and vacuum sweeping was carried out at RADC, Hanscom AFB.

All samples were cut with cross section of approximately $1.0 \times 1.0 \text{ cm}^2$ and thickness of approximately 1 mm, the latter dimension being parallel to the c -axis.

Irradiation and Measurement

Samples were irradiated at room temperature with X-rays from a tungsten target source operated at 20 mA and 40 kV. Soft X-rays were filtered out by a layer of sputtered silver electrode, and a glass filter approximately 1 mm thick. In addition, the NQ sample was also irradiated at room temperature with γ -ray from a Co source at Brookhaven National Laboratory, which has a dose rate of 1.1 MRöntgen/hr.

After irradiation, the samples were transferred to a Super Varitemp Cryostat (Janis Corp.), and cooled down to liquid helium temperature. An automated capacitance bridge (C. Andeen Assoc.) was used to carry out the dielectric loss measurements, which cover a frequency range of 30 to 100 kHz and a temperature range of 3 - 273 K.

Results and Discussion

Vacuum Swept Quartz Crystal

We would first like to address the question of the origin of the low temperature irradiation peak. As pointed out in the Introduction section, two different irradiation related defect centers have been suggested as the possible sources of this irradiation peak: (1) the Al-h center formed during irradiation, and (2) an alkali center that involves the Na⁺ ion released from Al-Na during irradiation. In order to investigate these possibilities, we have made several measurements on the vacuum swept crystal, H29-14 (see Table I). The crystal was first swept in air, and then in vacuum for an extended period of time to ensure that all alkali and hydrogen ions have been swept out.

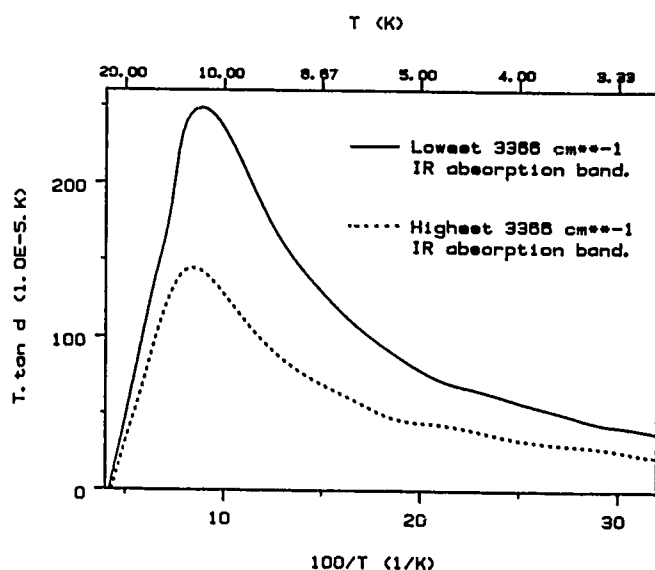


Fig. 2. Dielectric loss peaks, plotted as $T \cdot \tan \delta$ vs. $1/T$, for two vacuum-swept H29-14 samples with different 3366 cm^{-1} IR absorption: solid line, $\alpha_X = 0.025 \text{ cm}^{-1}$, and dashed line, $\alpha_X = 0.3 - 0.4 \text{ cm}^{-1}$.

The IR absorption spectrum of this vacuum-swept crystal measured by Lipson at RADC showed that the crystal is not uniformly swept. While the crystal has no 3581 cm^{-1} absorption (corresponding to grown-in OH) throughout, it showed varying degree of 3366 cm^{-1} absorption (corresponding to Al-OH) along the z-direction, i.e. along the c-axis. We studied two samples by means of dielectric measurements: one from the best swept region which has the lowest 3366 cm^{-1} absorption ($\alpha_X \sim 0.025 \text{ cm}^{-1}$), and one from the worst swept region which has higher 3366 cm^{-1} absorption ($\alpha_X \sim 0.3-0.4 \text{ cm}^{-1}$). The results of the measurements are shown in Fig. 2 as plots of $T \cdot \tan \delta$ vs. $100/T$. Despite the absence of alkali and the low level of hydrogen, both samples show a low temperature peak with maxima at $100/T \sim 8.5 \text{ K}^{-1}$. The height in the best swept sample is about twice that in the worst swept one, and their shapes are identical within experimental error. No observable change of the peaks occurs after a 321°C annealing of both samples.

It is interesting to compare the results of our dielectric measurements to the Al-h contents measured with ESR by Halliburton on samples prepared from the same vacuum swept crystal.[16] He measured 1.3 ppm of Al-h center in the best swept region, and 0.5 ppm in the worst swept region. The 2.6 : 1 ratio of the Al-h contents observed in this way correlates reasonably well with the 1.8 : 1 ratio of the dielectric peak heights that we observed in corresponding samples.

From the fact that both the "irradiation peak" and the Al-h center appear in these unirradiated vacuum-swept samples despite the absence of alkali, we conclude that the irradiation peak cannot be due to the alkali center, but that it must be some manifestation of the Al-h center.

Effect of Irradiation Dose

In order to investigate the effect of irradiation dose on defects in quartz crystals, we have carried out

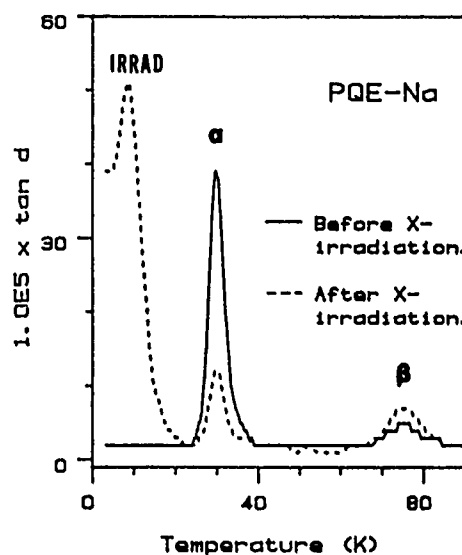


Fig. 3. Dielectric loss as a function of temperature before and after irradiation for a Na-swept PQE crystal as measured at 1 kHz. The data show the Al-Na α and β peaks as well as the irradiation peak.

a series of room temperature X-ray irradiations on a Na-swept PQE crystal, and measured the corresponding changes in the α and β Al-Na peaks as well as the irradiation peak. The various peaks are shown in Fig. 3, the results of the irradiation series are presented in Fig. 4. While the α peak decreases and the irradiation peak increases monotonically with the total dose, the β peak height goes through a maximum before it drops down to zero. By taking into account both the α and β peaks, however, the overall Al-Na concentration is found to be decreasing monotonically with the accumulating dose. The ratio R of α/β goes from an initial value of ~ 12 before the irradiation to a constant value of 2.7 ± 1 after irradiation. Values of R ranging from 10 to 17, depending on the previous heat treatment, have been observed previously,[4,5,8] but a value as low as 2.7 has never been reported. An even lower value of R of 0.85 was observed in this laboratory in a γ -irradiated, Na-swept NQ crystal.

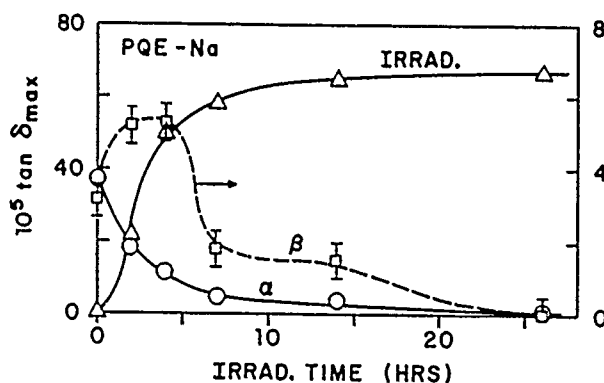
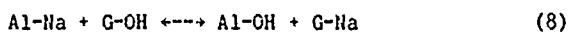
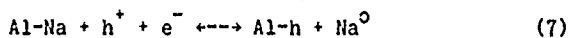
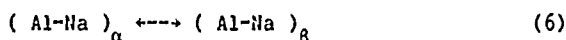


Fig. 4. Variation of the heights of the three dielectric loss peaks with X-ray irradiation time for a Na-swept PQE crystal. The left hand scale applies to the α and irradiation peaks, and the right hand scale to the β peak.

The R ratio, according to equation (5), is proportional to the concentration ratio of the nearest neighbor α to the next nearest neighbour β Al-Na centers. A typical value of $R = 13$ before irradiation gives $N_\alpha/N_\beta = 20 \pm 8.5$. The lower R values after irradiation correspond to smaller N_α/N_β ratios: $R = 2.7$ gives $N_\alpha/N_\beta = 4.2 \pm 1.8$, and the $R = 0.85$ mentioned above gives $N_\alpha/N_\beta = 1.4 \pm 0.6$. Thus it is reasonable to say that one of the effects of the irradiation is to redistribute the initial Al-Na population, which is in thermodynamic equilibrium between the α and β states, to a more random one in which the N_α/N_β ratio comes much closer to unity. The maximum in the β curve shows that initially there are more β centers formed than destroyed, presumably at the expense of the α centers.

Based on these considerations, we propose the following explanation for these observations. During irradiation, the Na⁺ ion, in an Al-Na center, whether α or β type, leaves the Al³⁺, presumably through the influence of the electrons and holes generated by the irradiation. The Na⁺ may then go to a yet unknown trap site, or return to an Al³⁺ to reform the Al-Na center. In such a recapture, the Na⁺ can go into either the α or β site. This randomizes the α/β population distribution, giving rise to a N_α/N_β ratio closer to unity, which results in a smaller R ratio. The maximum in the β curve is then the result of the combination of two effects: the randomization of the N_α/N_β ratio which causes the initial rise, and the decrease in the overall Al-Na concentration which causes the eventual drop in the β -curve.

This dynamic picture of the irradiation induced destruction and reformation of the Al-Na centers has an important implication: the dissociated Al³⁺ does not have to capture a Na⁺, but can instead capture a hole to form Al-h or a proton to form Al-OH center. The protons are presumably released from the grown-in OH centers, which we called G-OH, where G designates the yet unknown grown-in OH site. In other words, there can be these three competing processes occurring during irradiation: (1) the α/β population randomization, (2) the formation of Al-h, and (3) the formation of Al-OH. These three reactions can be represented with the following equations:



These processes, rather than occurring sequentially as previously thought, are occurring simultaneously, and are competing with each other throughout the irradiation process.

Effect of Annealing of an Irradiated Crystal

In order to study the annealing effects following irradiation, we utilized a Na-swept QA26 sample, which has a high Al content (see Table I). Following an initial 4 hr irradiation at room temperature, a series of isochronal (45 min) anneals were carried out, with results as shown in Fig. 5. The irradiation only lowered the α peak from 120 to 90 while producing an irradiation peak of height 65 (all $\times 10^{-5}$). The irradiation peak decreases strongly between 350 and 500 K and goes to zero at 600 K, while the α peak apparently increases in two stages, the larger one at ~ 650 K occurring after the irradiation peak is completely gone. The β peak (not shown in the figure

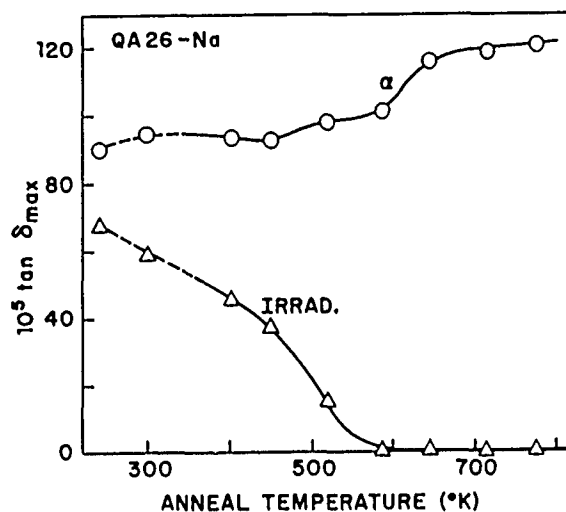


Fig. 5. Annealing of the α peak and the irradiation peak in a Na-swept QA26 crystal. The first points (shown below 300 K) were obtained immediately after irradiation; the second points (at 300 K) were measured after standing for 115 hr. All other points represent 45 min isochronal anneals.

for clarity) decreases so that the α/β ratio R returns immediately after the start of annealing to its pre-irradiated value of 13 ± 1 .

It is interesting to compare these results with the annealing behavior reported by Martin, who employed IR absorption to detect Al-OH centers and acoustic loss measurements to follow both an Al-Na loss peak and an irradiation induced loss peak at 23 K (for 5 MHz frequency) believed to be due to Al-h centers.[7] The Al-Na loss peak was found to be restored in two stages that match those observed in the present work. The first stage centered at 525 K, where the 23 K loss peak disappeared, and the second centered at 650 K where Al-OH anneals out.

These observations can be explained as follow: (1) prior to the occurrence of stage I, at just above room temperature, the thermodynamic equilibrium between α and β type Al-Na centers is restored. This causes the α/β ratio, R, to be restored to its pre-irradiated value. (2) Stage I annealing, which occurs between 450 and 550 K, involves the recombination and mutual annihilation of the electrons and holes, to eliminate the alkali centers and the Al-h centers, which cause the irradiation peak. This reaction is just the reverse of equation (7). (Additional support for the concept that electrons and holes become mobile in stage I comes from the fact that the E_1' center, which involves an electron trapped at an oxygen vacancy, only forms after irradiation followed by annealing at stage I temperatures.[18]) (3) Stage II annealing, which occurs between 600 and 700 K involves the re-exchange of Na⁺ and H⁺ to restore the G-OH and Al-Na centers, i.e., the reverse of equation (8). Thus, by stepwise reversal of each of the radiation produced reactions, eqs. (6)-(8), the crystal is finally restored to its as-grown condition.

Conclusions

We summarize in this section what has been learned in this work:

- (a) Both the Al-h centers and the "irradiation peak" are observed in the vacuum-swept quartz crystal. This shows that the irradiation peak is some manifestation of the Al-h centers.
- (b) During irradiation, the defect centers that give rise to the various dielectric loss peaks undergo dynamic destruction and reformation processes. Several reactions are competing with each other and occurring simultaneously, namely the Al-Na α/β population randomization, and the formation of Al-h and Al-OH centers at the expense of Al-Na centers.
- (c) During annealing, the various reactions that occurred during irradiation are reversed at successively higher temperatures:
 - The thermodynamic equilibrium between α and β Al-Na centers is restored at just above room temperature.
 - Above 450 K the Al-h centers are converted back to Al-Na, presumably with electron-hole recombination, giving rise to the stage I α -peak restoration.
 - Above 600 K the Al-OH centers are converted back to Al-Na giving rise to the stage II α -peak restoration.

While this picture of various dynamic processes occurring during irradiation and annealing of quartz crystals may be too simplified, and still leaves open the question of the nature of both the G-OH center and the trapping site for the alkali center, it nevertheless can serve as a starting point for a more detailed understanding of irradiation process in quartz.

Acknowledgement

This work was supported by the U.S. Air Force under contract F-19628-85-C-0104 and monitored by H.G. Lipson, RADC Hanscom AFB.

References

1. See numerous papers in Proceedings of the Annual Symposium on Frequency Control, 1978-1985.
2. B.R. Capone, A. Kahan, R.N. Brown and J.R. Buckmelter, IEEE Trans. Nucl. Sci., NS-13, 130 (1966); NS-17 217 (1970).
3. J.C. King and H.H. Sander, Radiation Effects, 26, 203 (1975).
4. J.M. Stevels and J. Volger, Philips Res. Rep., 17, 283 (1962).
5. D.S. Park and A.S. Nowick, Physica Stat. Solidi (a) 26, 617 (1974).
6. D.B. Fraser, in Physical Acoustics (ed. W.P. Mason), Vol. 5, Chap. 2, Academic Press, N.Y., 1968.
7. J.J. Martin, J. Appl. Phys. 56, 2536 (1984).
8. J. Toulouse and A.S. Nowick, J. Phys. Chem. Solids 46, 1285 (1985).
9. A. Kats, Philips Res. Rep. 17, 133 (1962).
10. R.N. Brown and A. Kahan, J. Phys. Chem. Solids 36, 467 (1975).
11. H.G. Lipson and A. Kahan, J. Appl. Phys. 58, 963 (1985).
12. M.E. Markes and L.E. Halliburton, J. Appl. Phys. 50, 8172 (1979).
13. L.E. Halliburton, N. Koumvakalis, M.E. Markes and J.J. Martin, J. Appl. Phys. 52, 3565 (1981).
14. W.J. de Vos and J. Volger, Physica 34, 272 (1967); 47, 13 (1970).
15. J. Toulouse and A.S. Nowick, in Proceedings of the Materials Research Society Symposium, Vol.24, p. 149 (ed. J.H. Crawford, Y. Chen and W.A. Sibley), Elsevier, N.Y., 1984.
16. L.E. Halliburton, private communication.
17. A.S. Nowick and W.R. Heller, Adv. Phys. 14, 101 (1965); 16, 1 (1967).
18. M.G. Jani, R.B. Bossoli, and L.E. Halliburton, Phys. Rev. 27, 2285 (1983).

BERLINITE : CHARACTERIZATION OF CRYSTALS WITH A LOW WATER CONCENTRATION
AND DESIGN OF BULK WAVE RESONATORS

J. Detaint¹, A. Zarka², B. Capelle², Y. Toudic³, J. Schwartzel¹,
E. Philippot⁴, J.C. Jumas⁴, A. Goiffon⁴, J.C. Doukhan⁵

¹ Centre National d'Etudes des Télécommunications PAB/BAG/MCT Bagneux 92220 - FRANCE.

² Laboratoire de Minéralogie et de Cristallographie Université de PARIS VI 75005 - FRANCE.

³ Centre National d'Etudes des Télécommunications LAB/ICM/MPA Lannion 22301 - FRANCE.

⁴ Laboratoire de Physico-Chimie des Matériaux Inorganiques - Université de Montpellier 34000 - FRANCE.

⁵ Laboratoire de Structures et Propriétés de l'Etat Solide Université de Lille 59655 - FRANCE.

Abstract

Several modifications to crystal growth methods were successfully investigated to decrease the water concentration in crystals. OH concentrations as low as 20ppm were measured by I.R. absorption, whereas resonators with enhanced Q factors were obtained from these crystals. Many Lang topographs were taken from samples grown in different conditions giving low water concentrations. They are compared to previously obtained ones. A study of the phase transition was performed, in Laue geometry with synchrotron radiation. The topographs obtained during this study with a "dry" crystal are very similar to the observations made for quartz crystals. An incommensurate phase having the same characteristics as for quartz was observed. In order to improve the design of berlinite resonators, an analysis of the thickness shear and thickness twist modes of plane Berlinite resonators with circular electrodes was implemented. This model is based upon the approximate equation established by H.F. Tiersten and a discretization of the continuity conditions at the electrode boundary. The principle and the advantages of the method of solution used are described. Results for several cuts of Berlinite are given and compared to similar results for quartz. Experimentally observed vibration modes of quartz and Berlinite are displayed and compared to computed ones.

Introduction

Berlinite belongs to the family of quartz analogous already mentioned in this symposium¹. Among all known analogous of quartz, berlinite is those which has the more similar physical and crystallographical properties to quartz and that have reached the most advanced state of development². The interest of berlinite results principally of enhanced piezoelectric and thermal properties^{3,4}. It was previously demonstrated that, this material presents cuts with zero frequency temperature coefficients either for bulk and surface waves^{5,6,7,8}. In recent years, the feasibility to grow large crystals was established⁹, and, it was demonstrated, as for quartz, that the water solubility in this material, could be noticeable and have a large influence on devices properties^{10,11,12}. More recently, a

theory of the incorporation of water, as point defects, was considered¹³ that have permitted to determine how to modify the growth parameters to reduce greatly the water concentration in $AlPO_4$ crystals¹².

In this paper a systematic investigation of the influence of growth conditions on water solubility in crystals will be reported together with results of several characterizations made on crystals with decreasing H_2O concentration.

Because of the fundamental importance of the nature of the $\alpha \rightarrow \beta$ phase transition on the thermal properties of devices, an in depth investigation of this transition was made and several important properties of the incommensurate phase existing between the α and the β phase of berlinite observed.

The last part of this paper will be devoted to a model established in view to optimize the design of energy trapping thickness shear resonators.

1. CRYSTAL GROWTH

To investigate the influence of the growth temperature on water solubility in crystals, growth experiments were conducted in the range 160-280°C by the slow heating method. The solvent was pure phosphoric acid at a concentration of 9.5 M. This concentration, much higher than the usual values^{14,16}, was chosen to increase the solubility of $AlPO_4$ at high temperature. The value of 70 % was retained for the filling of the autoclaves. Heating rates ranging from 1 to 4°C/day were used over temperature intervals generally close to 20°C.

The growth solutions were presaturated at the initial temperature for 48 hours to have the exact equilibrium concentration at this temperature¹⁵. To avoid any seed dissolution, these were put above the solution during heating to the initial temperature (Figure 1), then the autoclave was turned horizontal. In these conditions, the observed growth rate in the Y direction, were in the range .1 to .2 mm per day and per face, depending on the initial temperature.

Several experiments were done near 200°C at constant temperatures with varying filling to study the influence of pressure on water solubility in the crystals¹³. An horizontal autoclave was constructed for operation at high temperature in H_3PO_4 . The principle of this autoclave is displayed on figure 2.

HIGH TEMPERATURE SLOW HEATING GROWTH (PHOSPHORIC SOLUTION)

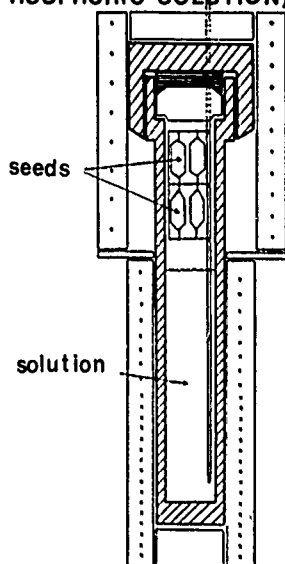


Figure 1 : High temperature growth.

COMPOSITE HORIZONTAL GRADIENT GROWTH

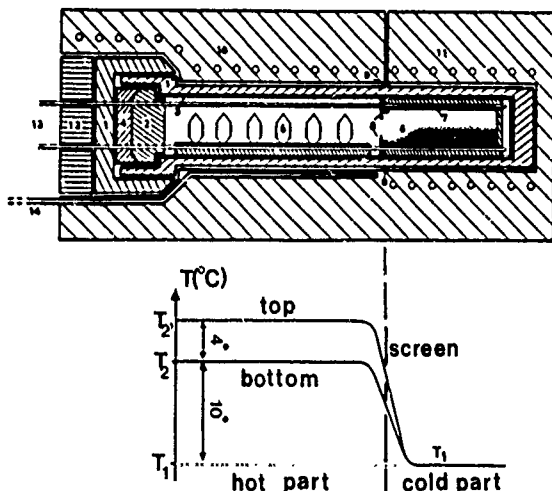
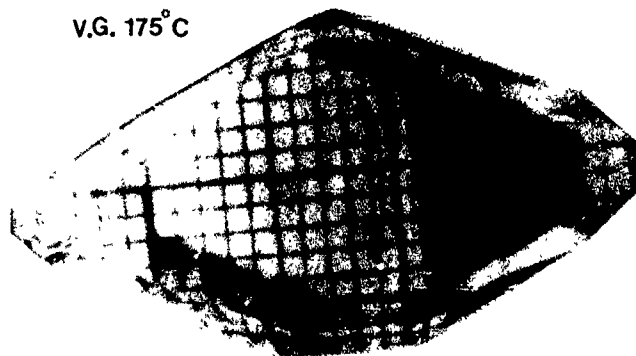


Figure 2 : Horizontal gradient Autoclave.

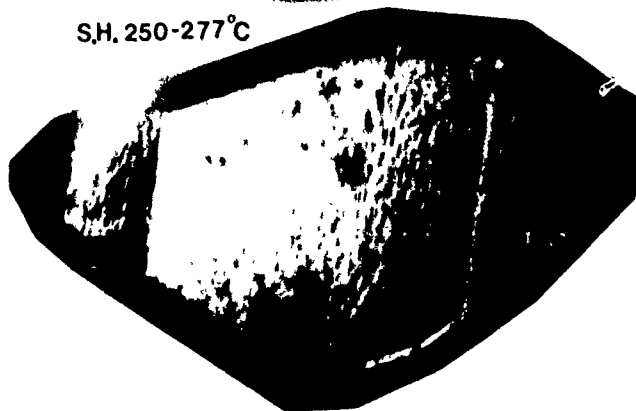
Three crystals obtained with Y seeds are compared on Figure 3. The first (at top of the figure) is typical of low temperature growth (vertical gradient near 175°C). The second (middle) was obtained by slow heating from 250 to 277°C. The etch figures that can be observed on the surface result of dissolution during cooling. The last crystal was obtained at high temperature with fast heating and fast cooling.

For most of the experiments the seeds were obtained from crystals grown at low temperature (160-190°C) by methods previously described. In several instances some water precipitation was observed to occur in these seeds, but with no negative incidence on the quality of the parts of crystals obtained at high temperature. On the contrary a significative improvement of the crystal quality was observed in these cases. Experiments with high temperature seeds were also successful.

V.G. 175°C



S.H. 250-277°C



V.G. H.T.

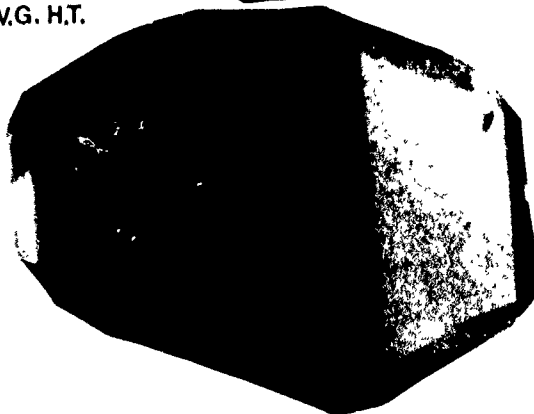


Figure 3 : Example of crystals.

II. INFRARED MEASUREMENTS

Infrared measurement were carried on Y plates cut outside the seed from the crystal obtained as previously reported. On figure 4 are displayed the absorbance of crystals grown at low to moderate temperature. The effect of growth temperature on water concentration is evident on this figure. Another effect can also be observed : the influence of pressure which was much greater in the case of vertical gradient (V.G).

On Figure 4 we can notice that the large concentration of water dissolved in these crystals produces a broad absorption band between 3600 and 2600 cm^{-1} with several superimposed peaks near 3500 and 3300 cm^{-1} .

For sample which has been heat-treated a reduction of the absorbance is observed. This results of the reduction of the concentration of water point defects that contribute to the absorbance in this wave number region to the benefit of water bubbles.

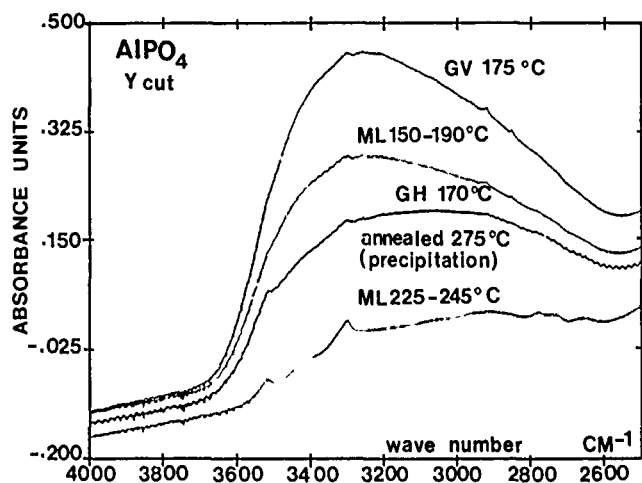


Figure 4 : Absorbance of water rich crystals.

On Figure 5 is given the absorbance of a crystal obtained between 250 and 277°C. On the curve with the very enlarged scale (at left), the peaks are well separated. As previously indicated by Steinberg, Chai et al¹⁰. The maximum at 3300 cm⁻¹ can be used to evaluate the water concentration in a manner similar to what is done for quartz with the absorbance at 3500 cm⁻¹.

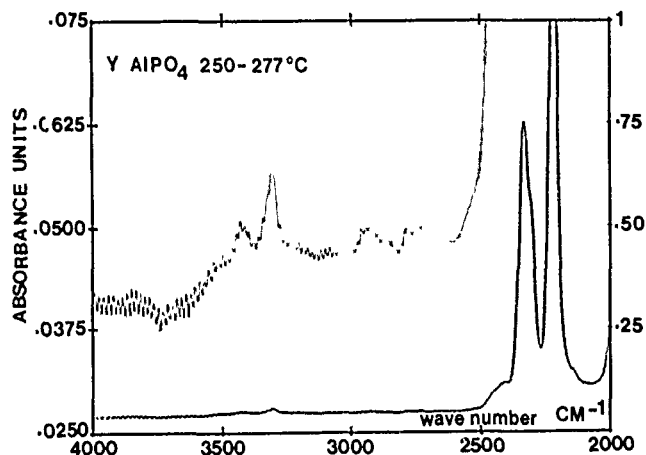


Figure 5 : Absorbance of a high temperature crystal.

On Figure 6, the absorbance of a Y plate of premium Q quartz is given for reference purpose. We must remark that the absolute intensities of the broad peaks related to the water concentration are not so different in the case of high temperature berlinite and good quartz.

The water concentration was calculated for many samples from absorbance data by the relation given in reference¹⁰. No corrections were made for the fact that the plates were Y cuts and not Z cuts and for the polarization, since the corrections are small. Some results are given in Table 1 where it can be seen that the solubility of water decreases as growth temperature increases; this solubility become very small above about 250°C and appears to follow approximately e^{-KT} variations as established by a theory¹⁷. Experiments made with different filling have revealed that the solubility of water decreases as the pressure in the autoclaves is lowered¹⁷.

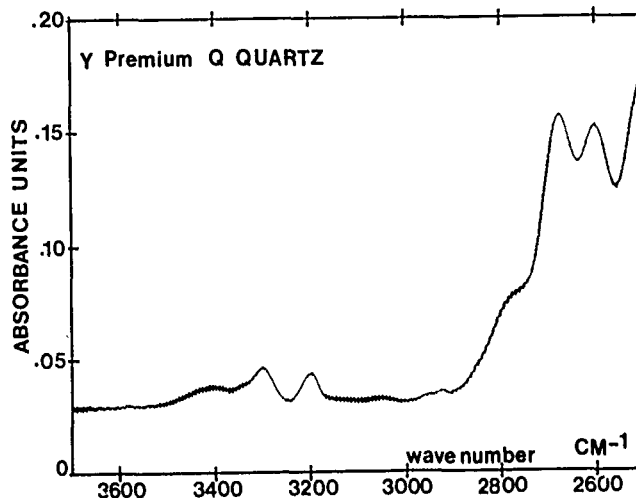


Figure 6 : Absorbance of a Y quartz plate.

Table 1 : Water concentration.

VARIATIONS OF [H₂O] WITH GROWTH CONDITIONS

CRYSTAL	α	[H ₂ O] ppm
GV 175° (HCL)	19.	- 900.
ML 150-190°	13.	- 600.
ML 150-190°	12.	- 680.
ML 210-225°	5.1	- 250.
ML 225-245°	1,66	- 80.
ML 225-245°	2.18	- 110.
ML 250-277°	0,65	- 30.
ML 250-277°	0,625	- 30.
best result high T° V.G.	0,380	- 18.

On Figure 7 is given the transmittance of an other berlinite plate obtained at 250°-277°C. This curve is to be compared in figure 8 with the transmittance of the Y plate of Premium Q quartz given in Figure 8. Again the comparison is not too unfavourable to berlinite even if we take in account the fact that the quartz plate is thicker by a factor of about 7. A very important property of Berlinite, also to be noticed on figure 7, is the much better transmittance than quartz in the range from visible to 4 μ m. This will probably lead to a lot of optical applications for Berlinite and AlPO₄ glass.

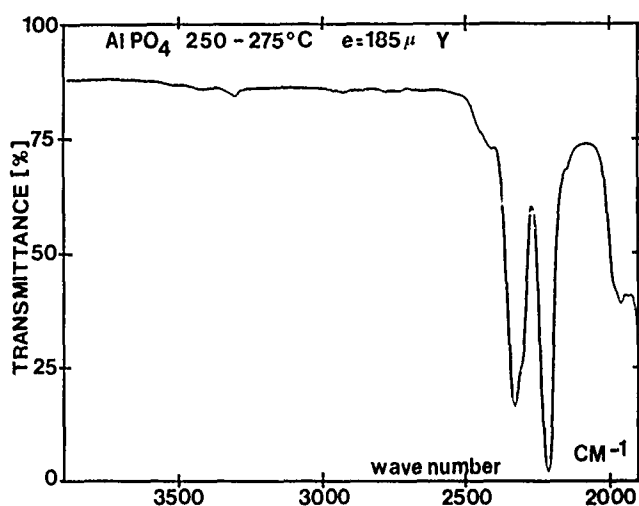


Figure 7 : Transmittance of a high temperature crystal.

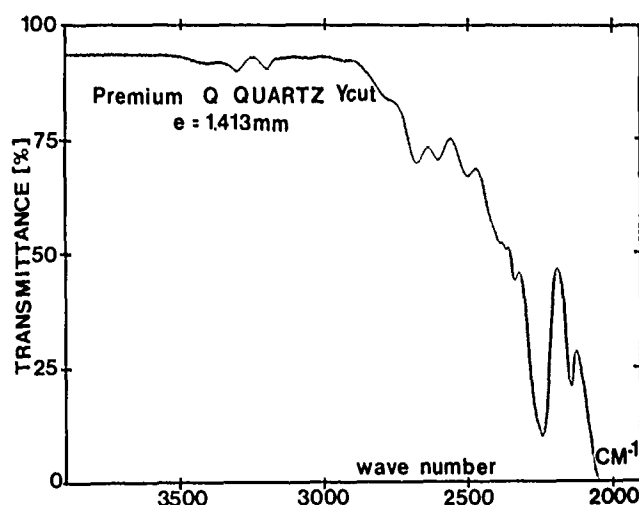


Figure 8 : Transmittance of a Y quartz plate.

III. ELECTRICAL MEASUREMENTS

Resonators were made from the plates cut outside the seed in crystals obtained at high temperature. Their frequency, frequency temperature coefficients, Q factors etc... were measured according to the standard methods established for quartz. On table 2 are displayed the results concerning the Q factors of plane resonators of Y cut operating at the fundamental mode. Assuming, as for quartz, that the product $Q \cdot f$ is a constant, the results are given at 5 MHz. These measurements were made on plates with thickness ranging from about .3 mm to .2 mm (about 6 to 9 MHz) and the devices Q factor obtained cannot be the intrinsic Q factor of the material which, as attested by some measurements made at the fifth overtone frequencies, are much higher (probably by a factor 10 for the best samples).

Table 2 : Q factor of Y cut Berlinite resonators.

INFLUENCE OF GROWTH T° ON Q FACTOR

CRYSTALS	Q factor* (5MHz)
VG.175°	$3.5 \cdot 10^3$
SH.203-215°	$6.9 \cdot 10^3$
SH.225-245°	$7.5 \cdot 10^3$
SH.250-277°	$65.4 \cdot 10^3$ [450. 10^3 5 th overtone]
SH.260-280°	$128. \cdot 10^3$

* measured with flat plates (f_{tal} mode), not intrinsic Q

During this study we have observed that the frequency constants of resonators, hence the stiffened elastic constant \bar{C}_{66} or $1/\rho$ of the material, increase as water is removed of the crystal. This confirms the observations reported last year¹² with the first high temperature crystals. We have also had the confirmation of the fact that the 2nd and 3rd order F.T.C have much smaller values with waterless berlinite and that they are lower than the corresponding values observed for similar cuts of quartz. The first order F.T.C of Y plates were found to be a little higher than one of the values given last year ; the value of $81. \pm 2.10^{-6} \text{°C}^{-1}$ was measured on more than 10 Y resonators cut in high temperature crystals. This value is still much lower than the value observed for Y cut of quartz. Similar observations were made for r or z cuts (Rhomboidal faces) so that our previous conclusions concerning the advantage of Y rotated cuts of berlinite over quartz¹² for thermal behaviour, are confirmed. In a similar manner we have had the confirmation that the coupling coefficients of berlinite resonators are most probably larger than the values that can be computed from the constants measured at the University of Maine⁸.

On Figure 9 is displayed the frequency temperature behaviour of Y resonators cut in a crystal obtained at high temperature on a low temperature seed. Three resonators cut outside the seed have a very similar quite linear thermal behaviour where as two resonators containing part of the seed display the typical behaviour of plates obtained at low temperature¹².

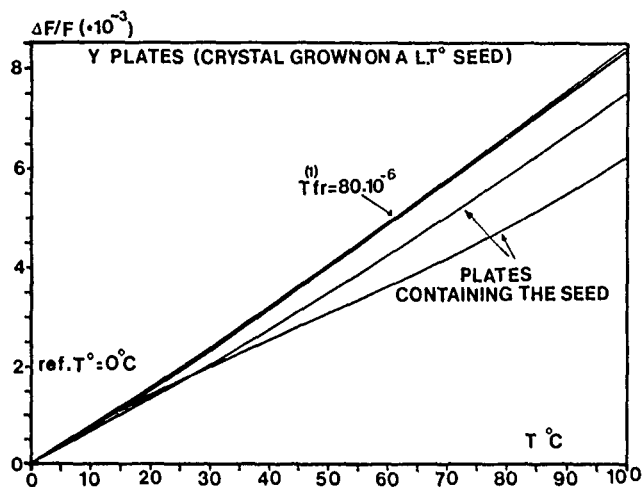


Figure 9 : $Fr(T)$ variations for Y cut Berlinite.

On Figure 10 are represented the $fr(T)$ variations measured with two Y and three r plates cuts outside the seed in another high temperature crystal. We can observe that the dispersion of the results are now very reduced ; the values of the first order F.T.C. for both cuts are displayed on this figure 10.

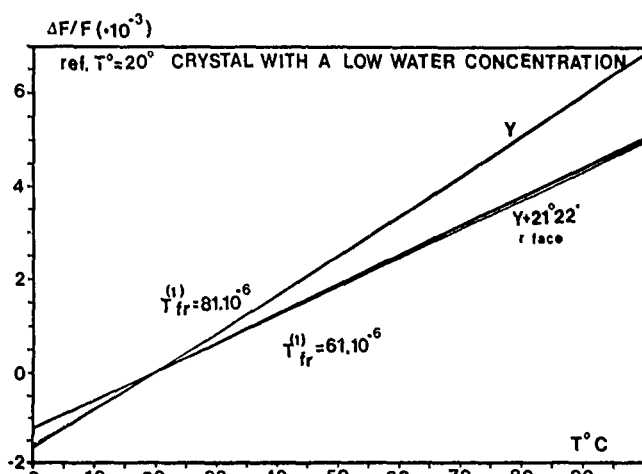


Figure 10 : $Fr(T)$ variations for Y and r cut of Berlinite.

IV - CHARACTERIZATION BY X-RAY TOPOGRAPHY

High temperature X-ray topographic studies were achieved using the ultra high vacuum system designed by Gastaldi, Jourdan and al¹⁸. This furnace associated with a TV camera and a video recording system allows to follow directly the α - β transition with the synchrotron radiation (LURE DCI, France) and the white beam topographic technique. Static topographs were also recorded on Kodak R-type film and Kodak ultra-speed dental film.

During this study, several differences were observed between crystals according to their water concentration. For instance in figure 11 a, a wet crystal showed some defects (dislocations D and growth bands B) at the beginning of the study at room temperature. On heating, a strong contrast appeared (Figure 11 b) and developed in the whole crystal up to the transition

temperature, so that the contrasts of the dislocations and growth bands were totally masked off. Another type of defect appeared after a temperature cycle around the transition (Figure 11 c) ; its density increased with the number of temperature cycles. At the end of the experience, only these defects and the strong contrast due to evolution of local strain were observed on the topograph. On the opposite for a dry crystal no contrast variation occurred.



11a



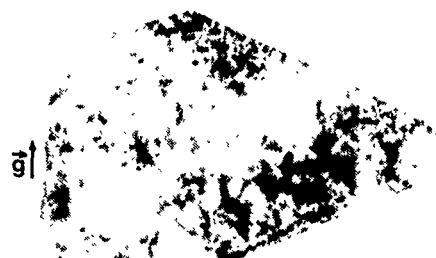
11b



11c

Figure 11 : Topographs of a "wet" crystal.

Figure 12 a presents a topograph obtained at room temperature of a dry berlinite Y-cut plate. The topographic contrasts show a low density of dislocation and a quite good quality of the whole crystal. The berlinite presents an incommensurate phase between the α and β phases and the phase front between the α and incommensurate phases can be seen in the central part of the sample in the topograph obtained at 953K (Figure 12 b).



12a



12b



12c

Figure 12 : Topographs of a "dry" crystal.

Study of the incommensurate phase

Due to the thermal gradient two zones are clearly discernable. The upper part corresponds to the incommensurate phase and the lower to the α phase. This topograph was obtained after a first temperature cycle around the transition and the two α_1 and α_2 twin domains appear in the α zone in white or black contrasts. It can be also noted in the incommensurate part, some contrasts which seem to extend from the twin boundaries present in the α zone. This type of defects has been

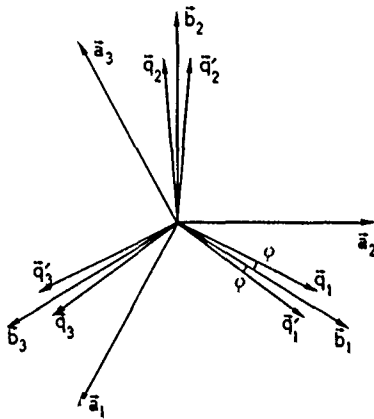


Figure 13

also observed in quartz crystal by different authors^{19,20}, and is probably at the origin of the twins. It disappears when the sample is totally in β phase (Figure 12 c).

The incommensurate phase presents a superposition of 3 modulation waves with 3 wave vectors q_i at 120° from each other (Figure 19). These wave vectors are tilted away from the $\langle 010 \rangle$ directions b_i of the hexagonal reciprocal lattice by an angle $\pm \phi$. The tilt induces the formation of $\pm \phi$ domains.

In a diffraction pattern the 3 modulation waves give 6 satellite spots around a Bragg spot, but this number becomes 12 with the splitting. This effect (Figure 14) has been observed by using a narrow beam and the white beam at LURE. We can see the Bragg spot (BS in figure 14) and 10 satellite spots. Each of the four spots close to the Bragg spot is splitted in two spots which present complementary contrasts : the $\pm \phi$ domains. The two other satellite spots show also a splitting with a domain structure but the two resulting spots overlap. This type of results was already obtained for quartz crystals but not yet for berlinite sample due to the poor quality of the crystals. The experimental evidence of the incommensurate phase and the $\pm \phi$ domains was allowed, in our case, by the good quality of the studied sample.

Figure 14

V - DESIGN OF BERLINITE PLANO-PLANO RESONATORS

This analysis was performed using the approximate equation for the slowly transversely varying thickness modes established by H.F. Tiersten first for monoclinic plates²¹, then for plates of arbitrary orientations^{22,23,24}. We have also used the approximate continuity conditions²¹ at the electrode edge consisting in the continuity of u_1 and $\partial u_1 / \partial n$.

For the electroded part of the plate the equation is with conventions and notations of references^{22,24,25} for Y rotated plates and for frequencies near the n^{th} overtone ($n = 1, 3, 5, \dots$) :

$$Mn \hat{u}_{1,11}^n + C_{55} \hat{u}_{1,33}^n - \frac{n^2 \pi^2}{4h^2} \bar{C}_{66} \hat{u}_1^n + \rho \omega^2 \hat{u}_1^n = \rho \omega^2 (-1)^{(n-1)/2} \frac{e_{26}}{C_{66}} 4 \frac{V_0 e^{i\omega t}}{n^2 \pi^2} \quad [1]$$

$$\text{where } \hat{u}_1^n = \hat{u}_1^n(x_1, x_3, t) \sin\left(\frac{n\pi x_2}{2h}\right)$$

For the unelectroded part we have :

$$Mn \hat{u}_{1,11}^n + C_{55} \hat{u}_{1,33}^n - \frac{n^2 \pi^2}{4h^2} \bar{C}_{66} \hat{u}_1^n + \rho \omega^2 \hat{u}_1^n = 0 \quad [2]$$

Eigen modes at $V = 0$:

The equations are normalised by division by C_{66} and put under the form of Helmholtz equations:

For the electroded part :

$$\frac{Mn}{C_{66}} \hat{u}_{1,11}^n + \frac{C_{55}}{C_{66}} \hat{u}_{1,33}^n + k_e^2 \hat{u}_1^n = 0 \quad [3]$$

For the unelectroded part :

$$\frac{Mn}{C_{66}} \hat{u}_{1,11}^n + \frac{C_{55}}{C_{66}} \hat{u}_{1,33}^n - k_l^2 \hat{u}_1^n = 0 \quad [4]$$

where :

$$k_e^2 = \frac{\pi^2}{4h^2} \frac{(f^2 - f_{ce}^2)}{f^2} \quad [5]$$

$$k_l^2 = \frac{\pi^2}{4h^2} \frac{(f_{cl}^2 - f^2)}{f^2} \quad [6]$$

$$f_{ce} = \frac{n}{4h} \sqrt{\frac{\bar{C}_{66}}{\rho}} \quad [7]$$

(cut off frequency for the electroded part)

$$f_{cl} = \frac{1}{4h} \sqrt{\frac{\bar{C}_{66}}{\rho}} \quad [8]$$

(cut off frequency for the unelectroded part)

$$f_{66} = \frac{1}{4h} \sqrt{\frac{C_{66}}{\rho}} \quad [9]$$

$$\alpha^2 = \frac{C_{66}}{Mn}, \beta^2 = \frac{C_{66}}{C_{55}} \quad [10]$$

Equations [3], [4] are transformed to isotropic Helmholtz equations by the transformation $x'_1 = \sqrt{C_{66}/Mn} x_1$, $x'_3 = \sqrt{C_{66}/C} x_3$; and then separated in polar coordinates r and τ ($r \cos \tau = x'_1 = \alpha x_1$; $r \sin \tau = x'_3 = \beta x_3$)

$$\hat{u}_1^n = R(r) \cdot T(\tau) \quad [11]$$

For the electroded part the resulting differential equations are :

$$\left\{ \begin{aligned} r^2 \frac{d^2 R}{dr^2} + r \frac{dR}{dr} + R(r^2 k_1^2 - v^2) &= 0 \\ \frac{d^2 T}{d\tau^2} + v^2 T &= 0 \end{aligned} \right. \quad [12]$$

$$\frac{d^2 T}{d\tau^2} + v^2 T = 0 \quad [13]$$

The separation constant v must be an integer since the solution T of [13] must be periodic with 2π as period : $v = m$

$$T = C_m^n \cos m\tau + D_m^n \sin m\tau \quad [14]$$

Equation [12] is a Bessel equations which has for solution bounded at $r = 0$ the Bessel functions of first kind²⁷ :

$$R = L_m^n J_m(r.k_e) \quad [15]$$

The general solution of [3] take the form :

$$\hat{u}_1^n = \sum_{m=0}^{m=\infty} J_m(r.k_e) \cdot (C_m^n \cos m\tau + D_m^n \sin m\tau) \quad [16]$$

Similary for the unelectroded part, the general solution can be expressed in function of the modified Bessel functions of second kind which are the solutions bounded at infinity of the differential equation obtained by separation of [4]

$$\hat{u}_1^n = \sum_{m=0}^{m=\infty} K_m(r.k_1) \cdot (E_m^n \cos m\tau + F_m^n \sin m\tau) \quad [17]$$

Generally the electrodes, and, the modes that can be excited in plates with parallel faces possess symmetries that permit to reduce expressions [16] and [17]. For example for electrodes symetric in x_1 and x_3 (round, rectangular etc...) only the symetric modes in x_1 and x_3 can be excited ; then the D_m^n and F_m^n vanishes. However, even in the case of symetric electrodes, it is often useful to consider also the antisymetric modes in x_1 or in x_3 that can be excited as a result of minor defects in practical resonators²⁸ (slightly unparallel plates, electrodes misalignment, electrodes tabs not exactly aligned in x_1 or x_3 direction etc... etc...). An example of such modes will be given below.

Boundary and continuity conditions

Due to the common dependance in $\sin(n\pi x_2/2h)$ of the solutions \hat{u}_{1e} and \hat{u}_{1l} , the boundary conditions on the major surfaces of the plates are satisfied to the requisited order of approximation²⁸. At all points of the electrode edge, we have to express the continuity of \hat{u} and of $\frac{\partial \hat{u}}{\partial n}$.

This was approximated by expressing these conditions at only a discrete number p of points (figure 15)^{29, 30}. For example, let us consider here after, the case of an electrode symetric, in x_1 and x_3 and an eigen modes of the same symmetry ($D_m^n = 0$; $F_m^n = 0$; $m = 2.i$). At the point $M^q(r^q, \tau^q)$ of the electrode edge the continuity conditions take the form :

$$\sum_0^\infty C_m^n J_m(r^q.k_e) \cos m\tau^q = \sum_0^\infty E_m^n K_m(r^q.k_1) \cos m\tau^q \quad [18]$$

$$\sum_0^\infty C_m^n r^q k_e J_m' \cos m\tau^q (\alpha N_1^q \cos \tau^q + \beta N_3^q \sin \tau^q) + \sum_0^\infty C_m^n m \cdot J_m \sin m\tau^q (\alpha N_1^q \sin \tau^q - \beta N_3^q \cos \tau^q) = \sum_0^\infty E_m^n r^q k_1 K_m' \cos m\tau^q (\alpha N_1^q \cos \tau^q + \beta N_3^q \sin \tau^q) + \sum_0^\infty E_m^n m \cdot K_m \sin m\tau^q (\alpha N_1^q \sin \tau^q - \beta N_3^q \cos \tau^q) \quad [19]$$

$$\sum_0^\infty E_m^n r^q k_1 K_m' \cos m\tau^q (\alpha N_1^q \cos \tau^q + \beta N_3^q \sin \tau^q) + \sum_0^\infty E_m^n m \cdot K_m \sin m\tau^q (\alpha N_1^q \sin \tau^q - \beta N_3^q \cos \tau^q)$$

where :

The index m has only even values $m = 2i$
 $\alpha = \sqrt{C_{66}/Mn}$; $\beta = \sqrt{C_{66}/C_{55}}$

$N_1^q(N^q, N^q)$ is the normal at point M^q

J_m, J_m', K_m, K_m' stand respectively for

$J_m(r^q.k_e), \frac{dJ_m(r^q.k_e)}{d(r^q.k_e)}, K_m(r^q.k_1)$ etc..

The series conditions in [18] and [19], expressed for the p points, are then truncated to p terms to obtain an homogeneous linear system, the determinant of which must vanish to have a non trivial solution in $C_0^n, \dots, C_{2p-2}^n, E_0^n, \dots, E_{2p-2}^n$. This constitutes the frequency equation to eigen modes at $V = 0$ that is numerically solved in the frequency interval $]f_{ce}, f_{c1}[$.

The eigen modes are obtained by solving the homogeneous linear system in $C_0^n, \dots, C_{2p-2}^n, E_0^n, \dots, E_{2p-2}^n$ at the corresponding eigen frequencies $f_{n\mu}$ ($\mu = 1, 2, 3, \dots$) in function of one of these coefficients.

Then, the eigen modes take the form :

$$\begin{cases} \hat{u}_{1e}^{n\mu} = \sin(n\pi x_2/2h) \hat{u}_{1e}^{n\mu} \\ \hat{u}_{1e}^{n\mu} = e^{j\omega t} \sum_{i=0}^{i=p-1} \mathcal{A}_{2i}^{n\mu} J_{2i}(r.k_e^{n\mu}) \cos(2i\tau) \end{cases} \quad [20]$$

for $M(r, \tau)$ in the electroded part and :

$$\begin{cases} \hat{u}_{1l}^{n\mu} = \sin(n\pi x_2/2h) \hat{u}_{1l}^{n\mu} \\ \hat{u}_{1l}^{n\mu} = e^{j\omega t} \sum_{i=0}^{i=p-1} \mathcal{B}_{2i}^{n\mu} K_{2i}(r.k_1^{n\mu}) \cos(2i\tau) \end{cases} \quad [21]$$

for the unelectroded part.

Where :

the \mathcal{A}_i and \mathcal{B}_i being of the form C_i/C_s and E_i/C_s

$$k_e^{n\mu} = \left[\frac{\pi^2}{4h^2} \left(\frac{f_{ce}^2}{f_{n\mu}^2} - f_{ce}^2 \right) / f_{ce}^2 \right]^{1/2}$$

$$k_1^{n\mu} = \left[\frac{\pi^2}{4h^2} \left(f_{c1}^2 - \frac{f_{n\mu}^2}{f_{c1}^2} \right) / f_{c1}^2 \right]^{1/2}$$

These eigen solutions must verify the orthogonality relation^{32, 33} :

$$\iiint_{Vt} \hat{u}_1^{n\mu} \cdot \hat{u}_1^{k\lambda} dV = \delta_{nk} \delta_{\mu\lambda} N^{(n)}(\mu) \quad [22]$$

with :

$$\iiint_{V_t} \hat{u}_1^{n\mu} \hat{u}_1^{k\lambda} dV = \iiint_{V_e} \hat{u}_{1e}^{n\mu} \hat{u}_{1e}^{k\lambda} dV + \iiint_{V_1} \hat{u}_{11}^{n\mu} \hat{u}_{11}^{k\lambda} dV$$

$$\begin{cases} V_t = \text{total volume of the plate} \\ V_e = \text{volume under the electrodes} \\ V_1 = V_t - V_e \end{cases}$$

The corresponding potential is :

$$\phi_e^{n\mu} = \frac{e}{\epsilon_{22}} \hat{u}_1^{n\mu} \left(\sin\left(\frac{n\pi x_2}{2h}\right) - (-1)^{(n-1)/2} \frac{x_2}{2h} \right) \quad [23]$$

Forced modes :

The steady state forced modes may be expanded in the form of linear combination of the eigen modes at $V = 0$ [23, 33].

$$u_1^* = \sum_n \sum_\mu H^{n\mu} \hat{u}_1^{n\mu} \sin(n\pi x_2/2h) \quad [24]$$

$$\phi^* = \sum_n \sum_\mu H^{n\mu} \phi^{n\mu} + \frac{x_2 V_0}{2h} e^{j\omega t} \quad [25]$$

where : $\hat{u}_1^{n\mu}$ is (as in [22]) $\hat{u}_{1e}^{n\mu}$ or $\hat{u}_{11}^{n\mu}$

$$u_{1e}^* = u_{1e} - \frac{e}{2h} \frac{x_2 V_0}{\epsilon_{22}} e^{j\omega t}$$

u_1 being the forced solution.

Accounting for the orthogonality of $\sin(n\pi x_2/2h)$ already expressed in [1] and [2]^{23, 25} ; we substitute the lateral dependance of u_1 i.e.

$$\hat{u}_1 = \sum_n \sum_\mu H^{n\mu} \hat{u}_1^{n\mu} \text{ in equations [1] and [2].}$$

Then we multiply the resulting expressions by $\hat{u}_1^{k\lambda}$ we integrate respectively on V_e and V_1 . The two expressions are then added, the result is :

$$\sum_n \sum_\mu H^{n\mu} \frac{\pi^2}{4h^2 f^2} (f^2 - f_{n\mu}^2) \iiint_{V_t} \hat{u}_1^{n\mu} \hat{u}_1^{k\lambda} dV$$

$$= (-1)^{(n-1)/2} \frac{\rho \omega^2 e}{C^2} \frac{4V_0}{n^2 \pi^2} e^{j\omega t} \iiint_{V_e} \hat{u}_1^{k\lambda} dV$$

using the orthogonality relation

$$\iiint_{V_t} \hat{u}_1^{n\mu} \hat{u}_1^{k\lambda} dV = 2h \iint_{S_t} \hat{u}_1^{n\mu} \hat{u}_1^{k\lambda} dS = 2 \delta_{nk} \delta_{\mu\lambda} N(n)(\mu)$$

We have :

$$H^{n\mu} = (-1)^{(n-1)/2} \frac{f^2 h e}{(f^2 - f_{n\mu}^2) n^2 \pi^2 C} \frac{4V_0 e^{j\omega t} \iint_{S_e} \hat{u}_1^{n\mu} dS}{N(n)(\mu)} \quad [26]$$

where : S_e = electroded surface
 S_t = total surface

The electrical response is obtained from the approximate electrical equation

$$D_2 = \frac{e}{\epsilon_{22}} u_{1,2} - \epsilon_{22} \psi_{,2}$$

$$D_2 = \frac{e}{2h} V_0 e^{j\omega t} (k_{26}^2 + 1) + \sum_n \sum_\mu (-1)^{(n-1)/2} H^{n\mu} \hat{u}_1^{n\mu} \frac{e}{h}$$

The admittance is :

$$Y = - \frac{1}{V} \iint_{S_e} D_2^* dS$$

$$Y = j\omega \left[\frac{S_e \epsilon_{22}}{2h} (k_{26}^2 + 1) - \sum_n \sum_\mu \frac{f^2}{(f^2 - f_{n\mu}^2)} \frac{4k_{26}^2 \epsilon_{22} h (\iint_{S_e} \hat{u}_1^{n\mu} dS)^2}{n^2 \pi^2 N(n)(\mu)} \right]$$

identifying with the admittance of the classical equivalent schema (one LC serie per mode) we found :

$$C_z = \frac{S_e \epsilon_{22}}{2h} (k_{26}^2 + 1) \text{ (Capacitance at } f = 0) \quad [28]$$

$$C_{n\mu} = \frac{f_2}{f_{n\mu}^2} \left(\frac{8k_{26}^2 \epsilon_{22}}{n^2 \pi^2 2h} \right) \frac{h (\iint_{S_e} \hat{u}_1^{n\mu} dS)^2}{N(n)(\mu)} \quad [29]$$

As expected from a distributed system, we find that " $C_{n\mu}$ " is frequency dependant. However $C_{n\mu}$ tends towards the equivalent motional capacitance of the mode as frequency approach $f_{n\mu}$. The rest of the infinite sum in Y contribute to the admittance for a small negative (capacitive) part.

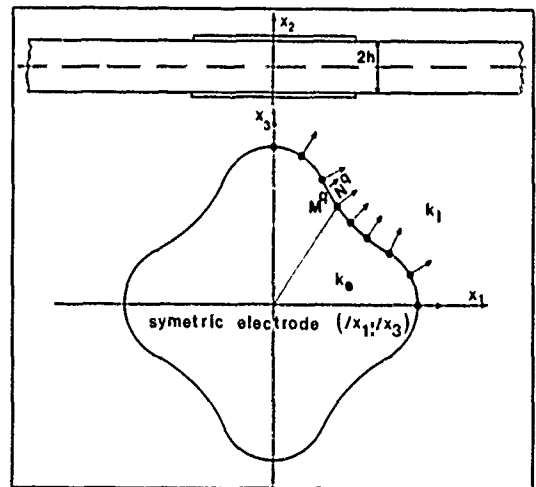


Figure 15 : Plate geometry and discretization.

Properties of the method of solution

A fast convergence is observed as the number of points where the continuity conditions are expressed is increased³⁰. As can be expected, from the properties of the solutions, for a given discretization the convergence is much faster for modes of nearly circular symmetry. This is the case for the 3rd overtone of AT quartz and also, to a least extend, for the first overtone of quartz and for the 3rd overtone of berlinite. This is also the case for some highly trapped anharmonics of the same modes when they possess nearly circular or radial nodal lines.

In figure 16 an example of the convergence properties is given concerning an AT quartz ($n = 1, \mu = 1$ and $\mu = 2, R = 1.92 \%$, circular electrode $r_e = 2,8 \text{ mm}$, $2h = 203 \mu\text{m}$). It can be noticed that an alternate convergence is obtained.

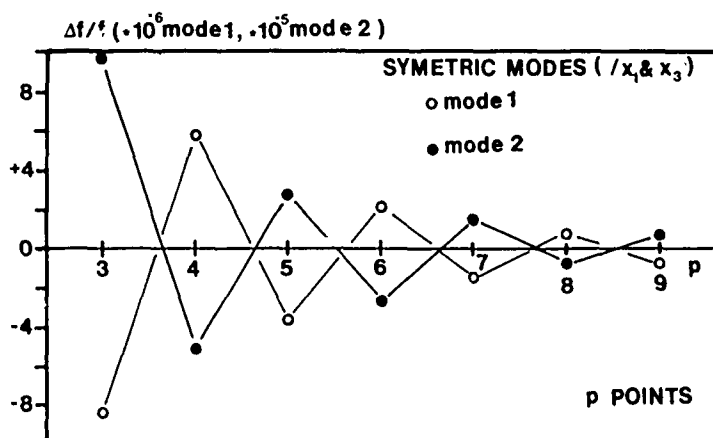


Figure 16 : Convergence of the solution.

This method of solution can consider various electrodes geometry²⁹. To illustrate this we have computed the modes of a resonator with hexagonal electrodes (AT quartz - same parameters as for figure 17 except $C = 2,8$ mm). The discretization and the fundamental mode ($n = 1, \mu = 1$) are displayed in figure 17 :

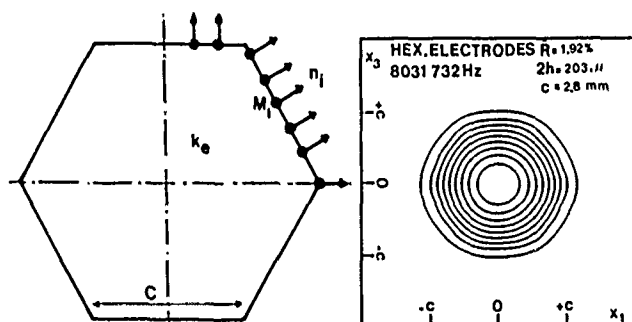


Figure 17 : Resonator with hexagonal electrodes.

It is also possible to take in account the effects on resonance frequencies and modes of the electrodes tabs³¹ in an approximate manner, provide they have the shape of circular sectors (figure 18). This is done simply by considering in the tab zones a solution of the form :

$$u_{1m} = \sum_{i=0}^{i=\infty} G_i K_i (r, k_m) \cos(i \tau)$$

$$\text{where: } k_m = \left(\frac{\pi^2}{4h^2} \frac{f_{cm}^2 - f^2}{f_{66}^2} \right)^{1/2}$$

f_{cm} being the cut-off frequency in the tab zones.

The continuity conditions are then expressed as previously described at the electrode edge. Doing this we neglect the continuity conditions along the radii limiting the tabs. However since the modes have most generably a nearly circular symmetry in the tab zones the continuity of the normal derivative is approximately verified, and since u_1 is small in these zones the continuity conditions concerning u_1 is of small importance³².

Although it is possible to correct this omission by perturbation analysis³⁴ it is probably sufficient to modify the discretization at the electrode edge in order to obtain a good approximation.

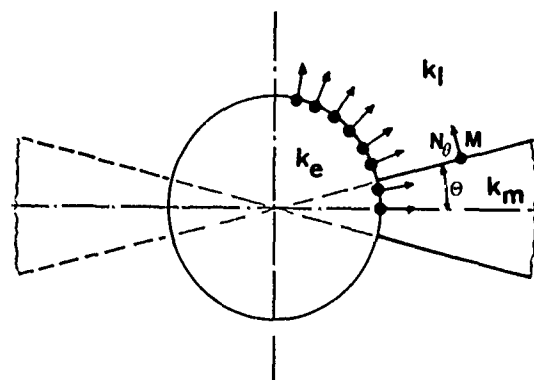


Figure 18 : Electrode tabs.

VI - COMPARISON OF BERLINITE AND QUARTZ RESONATORS

The model described in §V was used to compute the resonances frequencies and modes of resonators with circular electrodes made with several Y rotated plates of berlinite and quartz. The material constants and their temperature coefficients were obtained from the paper of Bailey et al⁸ for berlinite and of Bechmann³⁵ for quartz.

In table 3 are given the coefficients of equations [3] and [4] for the AT cut of Berlinite and quartz, for $n = 1, 3$ and 5, and the ratio M_n/C_{55} that characterize the lateral anisotropy. We can notice that AT berlinite is much more anisotropic than AT quartz.

Table 3a : Coefficients of the wave equation (in 10^{11} Nm^{-2})

	Mn(n=1)	Mn(n=3)	Mn(n=5)	C ₅₅	C ₆₆
Y-30° ALPO ₄	.8637	.5802	.7016	.5086	.2177
AT QUARTZ	1.099	.7580	.9009	.6880	.2924

Table 3b : Anisotropy factors $\sqrt{M_n/C_{55}}$

	n=1	n=3	n=5
Y-30° ALPO ₄	1.303	1.068	1.174
QUARTZ AT	1.263	1.049	1.144

To obtain a more precise comparison between the AT cut of quartz and cut Y-30° of berlinite we have computed the frequencies of the anharmonics of the first overtone for flat plates with circular electrodes as a function of $(r_e/2h)$ (r_e being the electrode radius). As suggested by the form of equations [3] and [4] and of the arguments of the Bessel functions, we have used as normalized frequency the expression

$$\left(\frac{f_{n\mu}^2 - f_{ce}^2}{f_{66}^2} \right)^{1/2} = \left(\frac{2h}{r_e} \right) \left(\frac{1}{\pi} \right) \left[r' \cdot k_{n\mu}^e \right]$$

We can observe on figure 19, that, with this representation, the diagrams are very similar for quartz and berlinite. However two important facts are to be noticed :

1°) Due to the lower values of the elastic constants of berlinite, to obtain the same frequency, it is necessary to use plates thinner by a factor of about 1.17, so that, a first corresponding reduction of r_e is required for this material.

2°) Since k_{26}^2 of berlinite is about twice that of quartz the frequency interval $f_{c1}-f_{ce}$ is greater for this material, so that, at equal $r_e/2h$ and equal R , it is possible to have more trapped anharmonics. This lead to a further reduction of r_e or \hat{R} .

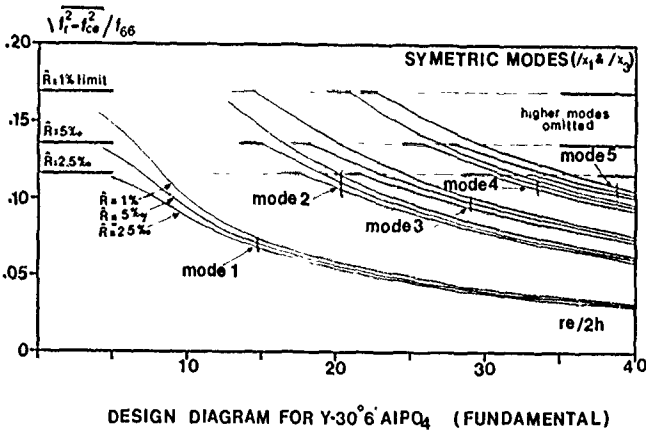
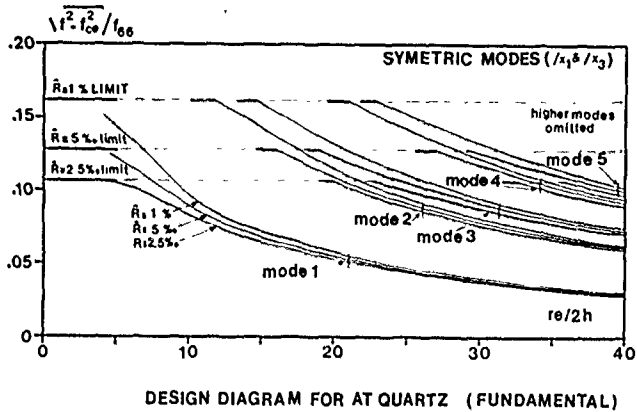


Figure 19 : Modes of resonators with circular electrodes.

VII - COMPARISON WITH EXPERIMENTAL RESULTS

The model was tested by comparing computed resonance frequencies and modes to experimental ones for quartz and berlinite.

AT quartz with circular electrodes

The first tests were made using quartz because of the much better precision on the material constants of this material. AT quartz plates of large diameter ($\phi = 16$ mm) a thickness of $203 \mu\text{m}$ and circular electrodes ($r_e = 2,8$ mm, $\hat{R} \# 1,92$ %) were used. The computed and measured frequencies are compared in Table 4 were it can be seen that a good agreement is obtained.

Table 4 : Comparison of measured and computed frequencies.

QUARTZ AT(Y-35°15') fundamental
2h=203 μ r_e =2.80mm \hat{R} =1.92 %

mode	Experimental	Computed
1	8027 152.	8026 748.
2	8101 824.	8109 275.
3	8140 596.	8143 546.

The electrical response is given in figure 20. Only two anharmonic modes are present with this design (large mass loading and medium electrode size).

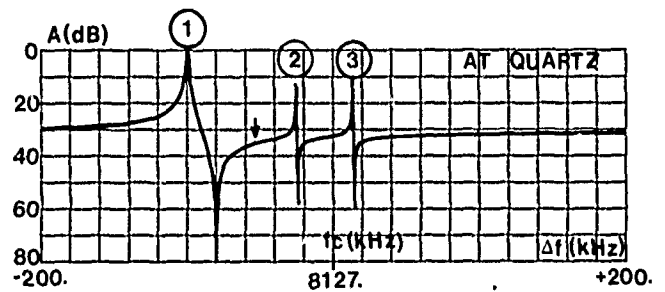


Figure 20 : Electrical response.

The corresponding computed eigen modes are displayed in figure 21. The mode are normalised to have a maximum value of 1.

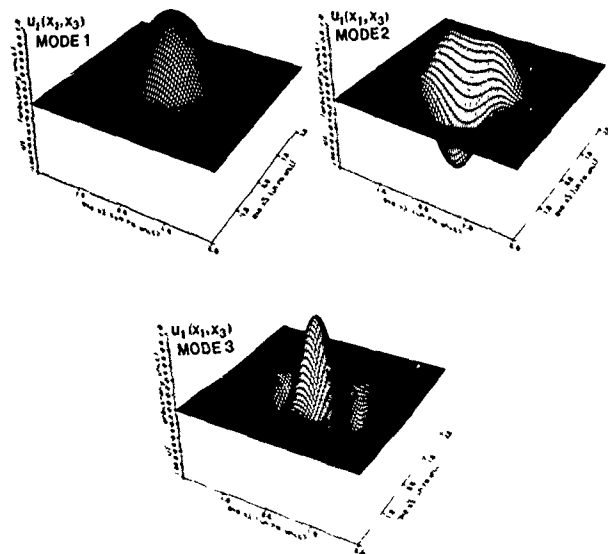


Figure 21 : Computed modes.

X-ray topography was used to observe and identify these modes, and, as it will be discussed below, we have chosen experimental conditions that permit to record on a film a picture representing with a good approximation $|u_1|$ for each mode. These topographs are compared with the computed values of u_1 represented as level curves, on figure 22.

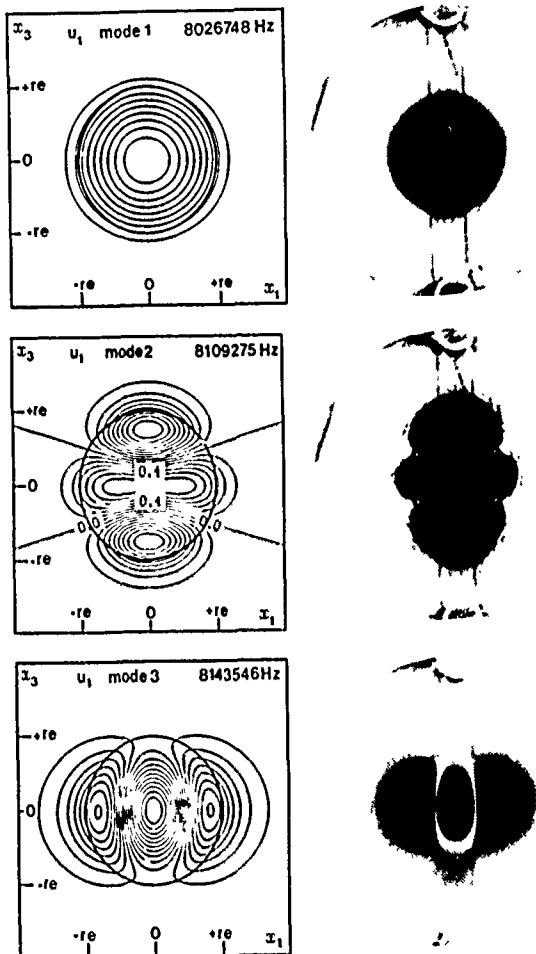


Figure 22 : Comparison of computed and observed modes.

Again, a good agreement can be noticed. The only minor differences that can be observed are tab effects (Not considered in this computation) of small importance for mode 1 and mode 2. For mode 3, a larger asymmetric tab effect is clearly visible ; it results, most probably, of a small defect in the parallelism of the plate.

This illustrates, the well known large sensitivities of the thickness modes of flat plates to minor technological defects (parallelism of plates, miss alignment of electrodes....). This sensitivity is increased in the case of large values of \hat{R}^{28} .

On X-ray topograph prints, it is difficult to make precision measurements, this can only be done by densitometry on the film or by directly measuring the diffracted intensity as a function of the lateral coordinates. The latter method was used to have simple quantitative verifications. On figure 23 we have represented the diffracted intensity and the resonator current as a function of the voltage at the first resonance frequency of a resonator of the previously described type. The diffracted intensity was measured at the center of the electrode. One can observe that for the low values of the voltage, the diffracted intensity is nearly proportional to voltage and current ; then a non

linear behaviour is observed that can result of either, the diffraction process, the counting process or, less probably, of mode non linearities.

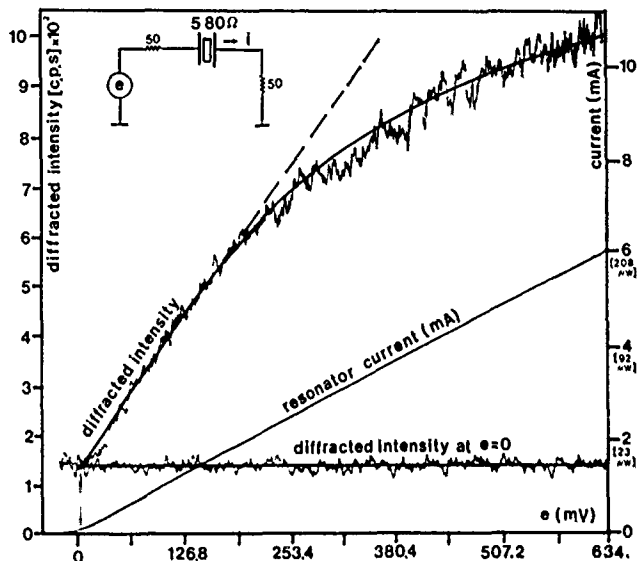


Figure 23 : Diffracted intensity as a function of excitation.

Using an excitation level within the quasilinear domain of figure 23, we have measured the diffracted intensity along the x_1 axis of the same resonator. The result is given in figure 24 where we have also represented the corresponding computed eigen mode, normalized to have the same maximum amplitude (at center of the resonator). The agreement between the two curve is evident.

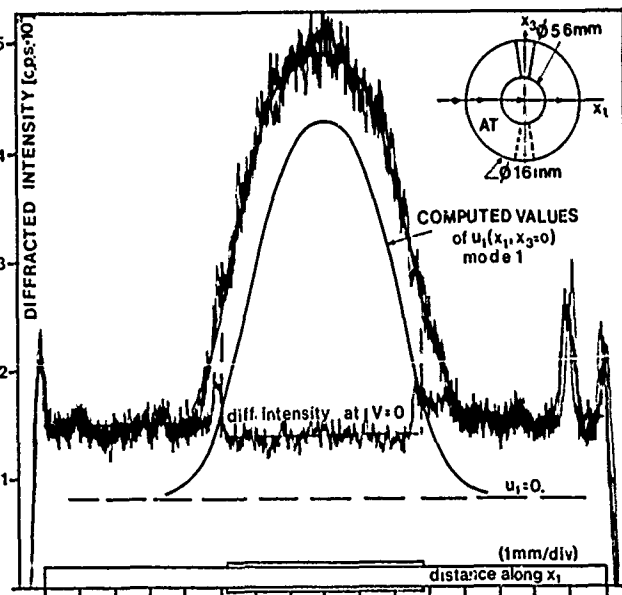


Figure 24 : Observed and computed modes.

Y cut of Berlinité

A slightly wedged plate ($2h = 213 \mu$ at the center $\pm 1 \mu$) was electroded ($r_e = 2 \text{ mm}$; $R = 1,28 \%$) and measured. The response curve is given on figure 25.

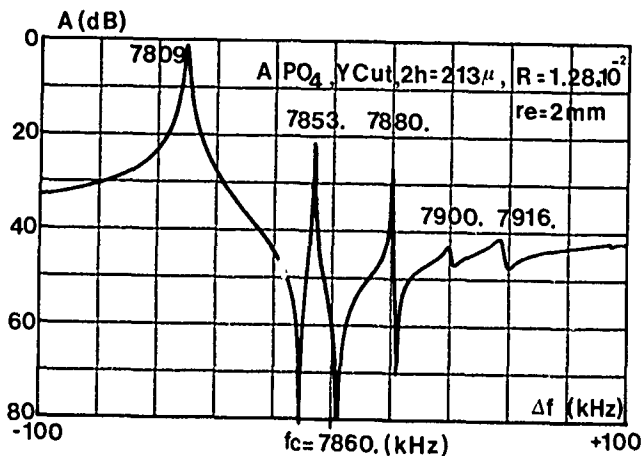


Figure 25 : Electrical response.

Five modes are experimentally observed, two of them being weak (not "well trapped"). As identified by X ray topography the 2nd and 3rd modes are respectively anti-symmetric in x_1 and in x_3 . Four computed mode only are found, two of them being the same antisymmetric modes. The experimental and computed frequencies are compared in table 5, a good agreement is found for the first three modes, a fair agreement is obtained for mode 4.

Table 5 : Comparison of computed and observed frequencies.

BERLINITE WEDGED PLATE, Y CUT

$$2h_c = 213 \mu\text{m} \quad re = 2\text{mm} \quad \hat{R} = 1.28\%$$

MODE	experiment.	computed
1	7809 (kHz)	7810.
2 AS./ x_1	7853.	7849.
3 AS./ x_3	7880.	7883
4	7900.	7897

The computed modes are represented on figure 26 in the form of $u_1^2(x_1, x_3)$.

The four first modes experimentally observed by X-ray topography are displayed in figure 27 where they are compared to computed ones. Again the agreement is very good for the first three modes. Mode 4, which is the first symmetric anharmonic with 2 nodal lines is greatly

affected by the defect of parallelism and the particular geometry of the plate, so that it has a non negligible amplitude at one of the mounting clips. The 5th mode experimentally observed results of the existence in the plate of a region with a higher f_{c1} cut-off frequency where it can propagate.

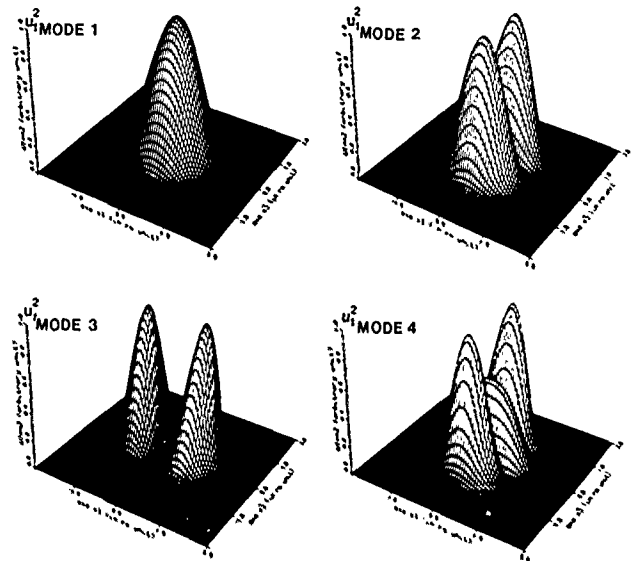


Figure 26 : Computed modes.

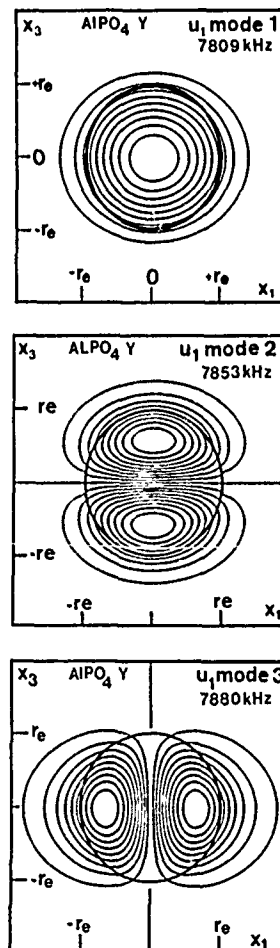


Figure 27 : Comparison of computed and observed modes.

A more rigorous analysis of modes of wedged plates will require to introduce in equations [3] and [4] the spatial variations of the thickness $h(x_1, x_3)$. This will lead to much more intricate solutions³⁶. However, this example, demonstrates that for a small defect of parallelism, the consideration of the antisymmetric eigen modes of the corresponding parallel plate can give sufficient information to identify these modes.

An important remark must be made again about the modes of berlinite: They are more anisotropic than the modes of the corresponding cuts of quartz. This is particularly evident for the first mode of figure 27 which is similar to what is observed for the Y-30° cut.

CONCLUSION

By growing crystals at different elevated temperatures we have demonstrated that the water solubility in Berlinite crystals decreases strongly when growth temperature is raised and pressure is reduced. This fact was already known for natural and synthetic quartz and predicted by the theory. With these crystals, we have demonstrated by I.R. spectrometry, that OH concentrations as low as in good quartz can be obtained in berlinite. Resonators made with low OH crystals display high Q factors and superior frequency temperature behaviour.

X-ray topography have indicated that improved crystalline quality are obtained with high temperature crystals. The study of the α β phase transition have demonstrated that this transition is totally similar to those of quartz. Several important properties of the incommensurate phase were found.

A model of plane resonators was implemented that permit the optimization of the design of berlinite resonators.

As the possibility to obtain large crystals has already been demonstrated⁹, the most relevant thing to do now is to find the growth conditions that permit the obtention of quantities of the large waterless crystals needed to produce performant SAW and bulk wave devices.

Acknowledgments: The authors wish to acknowledge the skilful assistance of C. Joly, R. Chenebault, A. Daniel, A. Jouanin, A. Jeanne Michaud, Y. Bernard for experimental work and measurements. The authors are indebted to M.H. Carru, Professor J.F. Petroff, Professor Maurin, and M.X. Buisson for valuable indications during the course of this work. They acknowledge helpful discussions with P. Tortelier and the expert assistance of M. J.Y. Moisan for I.R. measurements. Special thanks are due to P. Durand and M.P. Louis for preparing this manuscript.

Several part of this work were supported by CNET, DRET and AIVAR/CNRS contracts.

REFERENCES

- (1) J.H. Sherman Analogues of alphaquartz Proc. 40th Frequency Control Symposium this proc. (1986).
- (2) A. Goiffon, J.C. Jumas, M. Maurin, E. Philippot J. Solid State Chem vol 61 p 384 (1986).
- (3) Y.P. Chang, G.R. Barsch IEEE Su 23 n° 2 p 127 (1976).
- (4) A. Ballato, G.J. Iafrate Proc. 30th Frequency Control Symposium p 141 (1976).
- (5) P.H. Carr, R.M. O'Connell Proc. Frequency Control Symposium p 129 (1976).
- (6) J. D taint, M. Feldmann, J. H naff, M. Poignant, Y. Toudic Proc. 33rd Frequency Control Symposium p 70 (1979).
- (7) J. H naff, M. Feldmann, M.A. Kirov Ferroelectrics vol 42 pp 161-185 (1982).
- (8) D.S. Bailey, J.C. Andle, D.L. Lee, N. Soluch, J.F. Vetelino, BHT Chai Proc. 1983 Ultra Sonics Symposium p 335 (1983).
- (9) A. Ballato, E.R. Match, M. Mizan, B.H.T. Chai, R.S. Tilton, T.J. Lukaszec Proc. 38th Frequency Control Symposium (1984).
- (10) R.F. Steinberg, M.K. Roy, A.K. Estes, B.H.T. Chai Proc. 1984 IEEE Ultrasonics Symposium p 577.
- (11) J. D taint, M. Poignant, Y. Toudic Proc. 33rd Frequency Control Symposium p 93 (1980).
- (12) J. D taint, E. Philippot, J.C. Jumas, J. Schwartzel, A. Zarka, B. Capelle, J.C. Doukhan Proc. 39th Frequency Control Symposium p 234 (1985).
- (13) J.C. Doukhan, B. Boulogne, E. Philippot, J.C. Jumas, Y. Toudic Submitted to J. Crystal growth.
- (14) H. Poignant, H. Le Marechal, Y. Toudic Mat. res. bull vol 14 p 603 (1979).
- (15) E. Philipot, J.C. Jumas, A. Goiffon, R. Astier, M. Maurin Brevet Fran ais n° 2559165.
- (16) E.D. Kolb, R.A. Laudise J. Crystal Growth vol 43 p 313 (1978).
- (17) J.C. Doukhan, L. Trepied Bull Mineral. 108, 97 (1985).
- (18) J. Gastaldi, C. Jourdan, P. Mazo, C. Allasia, J.W. Jullier J. Appl. Cryst. vol 15 p 391-395 (1982).
- (19) K. Gouhara, N. Kato J. Phys. Soc. Jap. vol 54 p 1868-1881 (1985).
- (20) K. Gouhara, N. Kato J. Phys. Soc. Jap. vol 54 p 1882-1883 (1985).
- (21) H.F. Tiersten, R.C. Smythe J. Acoust. Soc. Am. vol 65 n° 6 p 1455-1460 (1979).
- (22) H.F. Tiersten, D.S. Stevens Proc. 36th Frequency Control Symposium p 37-46 (1982).
- (23) D.S. Stevens, H.F. Tiersten Proc. 39th Frequency Control Symposium p 436-447 (1985).
- (24) B.C. Peach Proc. 39th Frequency Control Symposium p 392 (1985).
- (25) See reference 17 of (22) or reference 16 of (23).
- (26) H.F. Tiersten J. Acoust. Soc. Am. vol 59 n° 4 p 879-888 (1976).
- (27) M. Abramowitz, H. Stegun Handbook of mathematical functions Dover NY 1972.

- (28) A. Glowinski, R. Lançon, R. Lefèvre
Proc. 27th Frequency Control p 233 1973.
- (29) K. Nakamura, H. Shimizu
Proc. 1976 Ultrasonics symposium p 606-609 (1976).
- (30) H. Sekimoto
IEEE trans. sonics. Ultrason. vol Su 31 n° 6 p 664-669
(1984).
- (31) H. Sekimoto, H. Nakata, M. Miura
Proc. 39th Frequency Control Symposium p 386 1985.
- (32) R. Holland, E.P. Eernisse
Design of resonant piezoelectric Devices M.I.T. Press
1969.
- (33) H.F. Tiersten
Linear Piezoelectric plate vibrations Plenum Press
(1969).
- (34) H.F. Tiersten, R.C. Smythe
J. Acoust. soc. Am vol 78 n° 5 p 1684-1689 (1985).
- (35) R. Bechmann, A. Ballato, T.J. Lukaszek
proc. IRE vol. 50 p 1818 1962.
- (36) B.K. Sinha, D.S. Stevens
J. Acoust. Soc. Am. vol 66 n° 1 p 192 (1979).

**The Influence of Surface Finish and
Metallization on Electrode Electromigration
in Alpha-Quartz During Sweeping**

John G. Gualtieri and Donald W. Eckart

US Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, New Jersey 07703-5000

ABSTRACT

We previously reported the temperature and dc-field values which maximize the migration of gold from electrodes into alpha-quartz during sweeping. It was suggested that the agglomeration of colloidal particles of electrode metal at dislocation and growth-tunnel sites was effective in reducing the density of etch channels in Au-swept quartz to zero.¹ However, an older report, concerning the anelasticity of quartz, showed how the migration of gold from electrodes into quartz bars was responsible for great increases in internal loss during longitudinal oscillation of the bars.² This report of internal dissipation, caused by the presence of gold in quartz, was the motivation for the investigation of a number of conditions to further understand electrode electromigration and its suppression in alpha-quartz. The conditions included: 1) electrode surface finish, 2) electrode metallization type, 3) use of diffusion barrier metal and refractory films, and 4) quartz of high and low purity. Other sweeping experiments contributed additional information on: 1) the agglomeration of colloidal metal particles along c-axis channels and along similar channels, 2) the electrodiffusion of impurities to the cathode surface forming circular deposits, and 3) the surface electromigration of molybdenum when gold-plated molybdenum ribbon is used for electrical contact.

INTRODUCTION

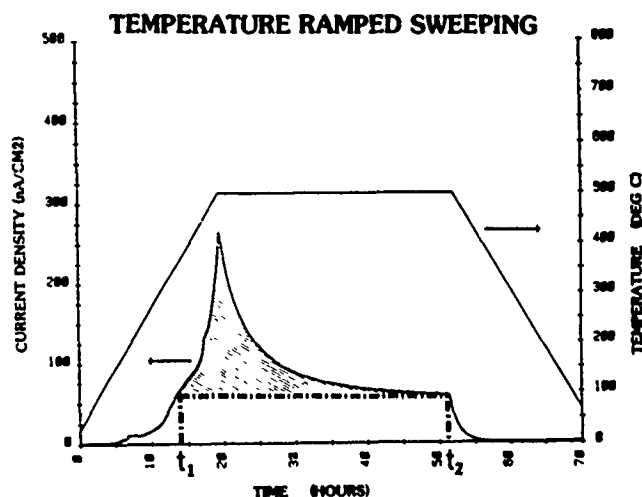
Sweeping is the currently used method for the elimination of alkali ions and other impurities incorporated during the growth of quartz. The elimination of these impurities is known to reduce the radiation sensitivity of resonators and to lower the etch-channel density of resonator blanks. Usually employed are Pt-foil electrodes under pressure, or evaporated or sputtered metallizations with pressure or welded contacts. It was previously reported that Au, Pt, Al, and Cu did not migrate into quartz when metal foils of these elements were used to sweep quartz.³ We found that considerable electromigration of electrode material was occurring during sweeping when evaporated, sputtered, or metal foil electrodes were employed.¹

If, for example, gold diffuses at resonator processing temperatures, the aging of resonators using gold electrodes may be adversely affected. It was previously reported that gold diffuses into quartz, without the aid of an electric field, at temperatures as low as 200°C.^{4,5} Electrode metal impurities have been shown to be coupled to ultrasonic vibrations in quartz, leading to great increases in internal loss.^{2,6} It was further suggested that diffusion of these impurities was influenced by the gradient of stresses generated during vibration.⁶ The mass transport of these metal impurities during resonator operation could change the resonator mass distribution in the long term, contributing to aging. Brice reported that surface damage influences the diffusion of electrode material into quartz at annealing temperatures between 200 and 400°C.⁴ These reports prompted our investigation. Our goal was to improve the technology of sweeping by developing methods to suppress electromigration of the electrode material.

EXPERIMENTAL

The cultured quartz used in this investigation was obtained from U.S. and non-U.S. suppliers. The quartz was lumbered into Y-bars or SC-sticks⁷ after seed removal. Evaporated Cr/Au and TiB₂/Au electrode metallizations were applied directly to the differently finished Z-surfaces.

Electrodes consisting of Pt-foil and depositions of carbon and boron nitride between the quartz and the Pt-foil were also employed. Gold-plated molybdenum leads were either welded or pressed against the metal electrodes. All sweeping was performed in air. Temperature ramped sweeping was employed; a description of the procedures was given previously.¹ At the completion of each run, a computer generated plot of current density vs time (see Fig. 1) was recorded and the charge transported per unit area of electrode was computed. Sample surfaces were surface profiled using a Talysurf profilometer, and were examined using a scanning electronic microscope (SEM) with energy dispersive analysis capability. Internal characteristics were studied using dark-background light-scattering microscopy. Light sources were either a high-intensity, fiber-guided, narrow beam illuminator or a low-power HeNe laser.



A typical computer plot of the current density and temperature vs time for a sample which was swept using evaporated Cr/Au electrodes. The transported charge per unit area is the integral of the current density above the final (stabilized) current density (shaded area) over time.

Figure 1.

EXPERIMENTAL OBSERVATIONS

In order to find the dependence of the bulk electromigration of gold on surface finish, a series of 2 x 2 cm Z-cut samples 3 mm thick were cut from a single quartz bar and prepared with different surface finishes before electroding with 10nm Cr, 5nm Cr/Au, and 100nm Au. The samples were then subjected to temperature-ramped sweeping. The results are given in Table I. The transported charge per unit area of electrode (Q/a) is the integral of the current density (J) above the final (stabilized) current density (J_s) over time. Referring to Figure 1

$$Q/a = \int_{t_1}^{t_2} (J - J_s) dt.$$

There is a possibility that the current density between t_1 and t_2 is partially due to conduction of grown-in impurities such as alkali ions or protons. However, if one compares successive sample slices from a single quartz bar, this should largely discount the contribution of grown-in impurities to the current density. Thus, we attribute differences in transported charge between "identical" samples to be related to the amount of electromigration of electrode material.

In all sweeping runs, the activation energy during cooldown was determined by plotting the logarithm of the (conductivity x temperature) vs reciprocal temperature. The activation

energy is related to the slope of the resulting straight line. Activation energies of 1.64 ± 0.07 eV were found regardless of surface finish or type of metallization. There is very good agreement with the value of 1.61 eV for steady axial conduction in quartz.⁸

The steady-state conduction in quartz is thought to be due to the transport of protons through the lattice. We interpreted the good agreement as an indication that the sweeping of grown-in impurities and electrode material was complete.

In Table I, we list the transported charge for each surface finish for which we have sweeping data. For the lapped (12 μ m) treatment we used 12 μ m aluminum oxide MICROGRIT abrasive.⁹ The polished (1 μ m) treatment was completed using 1 μ m cerium oxide MIROMET polishing compound.¹⁰ Roughness and depth of damage estimates were obtained, and will be discussed later.

TABLE I

SURFACE FINISH vs. TRANSPORTED CHARGE

SURFACE TREATMENT	ROUGHNESS RMS z-deviations (μ m)		DEPTH OF DAMAGE est. (μ m)	Q/a (coul/cm ²)
	INITIAL	FINAL		
Sawn (120 grit)	1.96	0.63	90	0.136
Lapped (12 μ m)	0.49	0.14	45	0.078
Polished (1 μ m)	0.021	0.013	21	0.037
Chemically polished	0.012	0.012	none	0.018

Samples were 2 x 2 x 0.3 cm, z-cut slabs cut from the same quartz bar. Metallization was Cr/CrAu/Au, sweeping field was 1000 V/cm. Samples were etched in $\text{NH}_4\text{F}:\text{HF} = 1:1$ at 75°C. Etch rate for z-outs $\approx 3 \mu\text{m}/\text{min}$.

Table II compares the transported charge for a number of electrode types. The carbon electroding was performed by coating the quartz Z-surfaces with a suspension of conducting carbon particles in amyl acetate and evaporating to dryness. The TiB_2/Au electroding was accomplished by evaporating 50nm TiB_2 compound and then evaporation of 100nm Au.

TABLE II

TRANSPORTED CHARGE vs. ELECTRODE TYPE

ELECTRODE TYPE	SURFACE FINISH	Q/a (coul/cm ²)
C/Pt-foil	Polished (1 μ m)	0.188
Cr/CrAu/Au	Polished (1 μ m)	0.126
Pt-sput/Pt-foil	Polished (1 μ m)	0.056
TiB_2/AuPt -foil	Polished (1 μ m)	0.043
Pt-foil	Polished (1 μ m)	0.013

All samples were cut from the same quartz bar (designated CULTURED #1, see Table IV). Sweeping field was 1000 V/cm.

Table III is a list of attempts made to limit the electromigration by using

diffusion barriers. The boron nitride coating was deposited by applying Boron Nitride Lubriccoat,¹¹ a suspension of insulating particles in a water solution and then evaporating to dryness.

TABLE III

SUPPRESSION ATTEMPTS vs. TRANSPORTED CHARGE

SAMPLE	ELECTRODE TYPE	SURFACE FINISH	Q/a (coul/cm ²)
A	Cr/CrAu/Au	Polished (1 μ m)	0.126
A	TiB ₂ /Au	Polished (1 μ m)	0.043
B	Cr/CrAu/Au	Polished (1 μ m)	0.020
B	TiB ₂ /Au	Polished (1 μ m)	0.013
C	Cr/CrAu/Au	Sawn (120 grit)	0.136
C	TiB ₂ /Au	Sawn (120 grit)	0.042
D	Pt-foil	Sawn (120 grit)	0.017
D	BN/Pt-foil	Sawn (120 grit)	0.009

Table IV compares quartz material of different suppliers. Samples were swept with Cr/Au electrodes on polished z-surfaces.

TABLE IV

TYPE OF QUARTZ vs. TRANSPORTED CHARGE

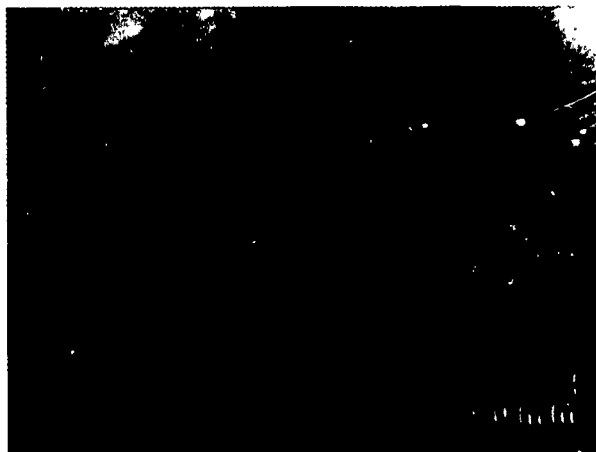
SAMPLE	ALPHA* (cm ⁻¹)	Qir* x 10 ⁶	Q/a (coul/cm ²)
Natural	0.165	1.02	0.194
Cultured #1	0.094	1.80	0.126
Cultured #2	0.072	2.35	0.046
Cultured #3	0.061	2.77	0.041
Cultured #4	0.058	2.91	0.041
Cultured #5	0.057	2.98	0.024

Samples were swept using Cr/CrAu/Au electrodes on polished (1 μ m) surfaces.
alpha* and Qir* were evaluated at 3410 cm⁻¹ using relationships suggested by Brice and Cole in Proc 32nd AFCS (1978) p. 1

Using dark-background light-scattering microscopy, we have observed the following internal features: 1) large agglomerations of colloidal metal particles, which frequently cause internal fracturing in the vicinity of the cathode; 2) colloidal metal particles decorating dislocations, planar growth bands, c-axis channels, and channels nearly perpendicular to external z- and r-faces. (See Fig. 2 a, b, and c.) We have also observed isolated deposits on the cathode surface after peeling away the Cr/Au metallization. Using energy dispersive analysis in a scanning electron microscope the deposits were identified as Na, K, Ca, Mg, Al, and Ti. (See Fig. 3.) Finally, if one uses gold-plated molybdenum ribbon as lead-in wire, care should be taken to suppress the possible surface electromigration of Mo. Sometimes the surface electromigration is extensive enough to move Mo down the side of the sample, causing intermittent electrical shorts. Roughening the sides of the sample and moving the Au/Mo lead away from the sample edge eliminates this problem.



a)



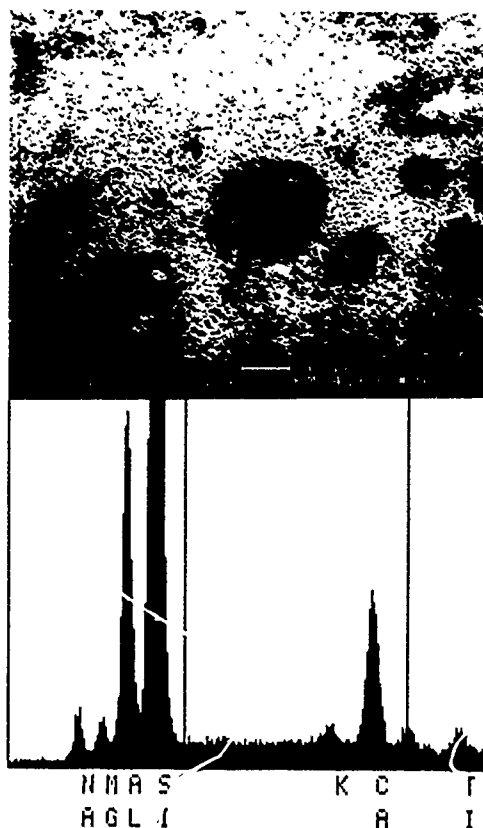
b)



c)

Light-scattering photomicrographs of swept samples. The light regions are gray planar bands a) using Pt-foil pressure electrodes, bands oriented parallel to the c-axis, 7.5X, b) using Pt-sputtered electrodes, bands oriented nearly perpendicular to external r-faces, inclined 70°C to the c-axis, 22X and, c) using Pt-foil sputtered electrodes, bands oriented nearly perpendicular to external z-faces, inclined 59°C to the c-axis, 22X.

Figure 2.



Upper-portion - Secondary electron image of cathode area of a swept quartz bar after peel-back of gold electrode. The spots are isolated deposits of impurities. Lower portion - Energy dispersive x-ray spectrum of a typical isolated deposit.

Figure 3.

DISCUSSION

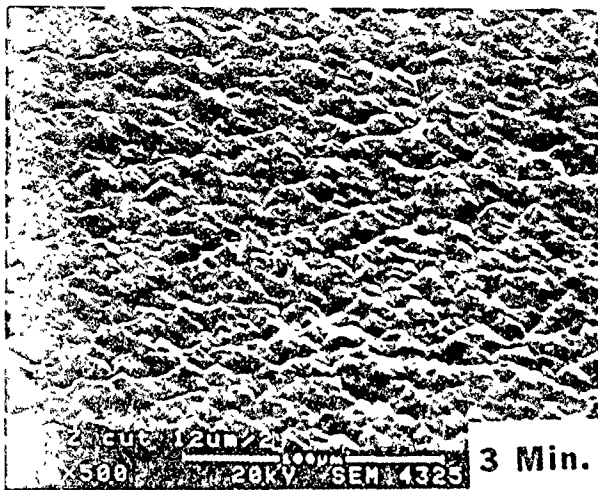
Brice reported that Au, Ag, and Cu electrode materials migrate into AT-cut quartz at annealing temperatures of 200-400°C and the diffusion is influenced by surface damage caused by lapping or polishing.⁴ From coarse grinding experiments on quartz spheres, we expect Z-faces to contain microscopic cleavages parallel to the original m, r, and z-faces.¹² Thus, from the penetrant angles of the microscopic cleavages, we expect Z-surfaces to be more deeply damaged than AT-cut surfaces, which are nearly parallel to the minor rhomb z-faces. In sweeping experiments, the electroded surface is perpendicular to Z, and the presence of a high dc field assists the in-diffusion.

In order to estimate the depth of damage produced by each surface treatment, companion (similarly treated) samples were etched together in a 1:1 solution of NH₄F:HF at 75°C for 5-minute periods. After each etch the samples were removed, surface profiled using a Talysurf profilometer and examined with an SEM.

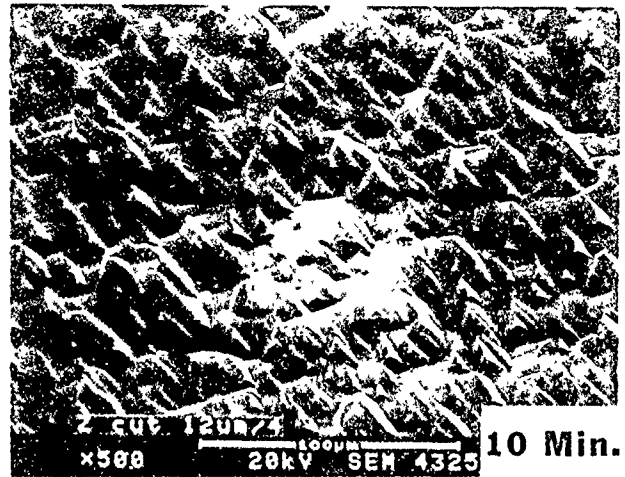
In Fig. 4, we show a composite of SEM photographs (taken after the indicated etching times) of a sample lapped with 12 μm aluminum oxide abrasive. Talysurf scans were made only in the X-direction, even though we found no dependence of roughness on direction. From Fig. 4 we see that, as etching progresses, triangular areas begin to appear. At this point, most of the damage has been removed. Further etching leads to very little change in roughness. Typical Talysurf scans are shown in Fig. 5. The scans were digitized and the method of least squares was used to fit the data to a straight line. The root-mean-square of the Z-deviations was taken as a measure of the roughness. In Fig. 6 a) and b) we show plots of surface roughness vs etching time for the indicated surface treatments. In each case, there was a marked decrease in surface roughness after a particular etching time and, then, no more reduction in roughness. The damage depth was estimated by multiplying the etching time, when roughness was reduced, by the etch rate of ~3 μm/min for Z-cuts.¹³ The damage depths for Z-cuts are deeper than those found for similar treatments using AT-cuts.^{4,14} The results shown in Table I indicate that the roughness and depth of damage produced by each surface treatment correlate well with the transported charge.

If we rank the electrode types we investigated according to transported charge, as in Table II, we find the Pt-foil electrodes have the lowest transported charge. However, we can see that a promising candidate for improved suppression of electromigration is TiB₂/Au on mechanically polished surfaces.

Our attempts to suppress electromigration by using diffusion barriers are shown in Table III. In each case, there is a reduction when the barrier is used. However, we found variability both in the electromigration and in its suppression when comparing samples from different suppliers. Table IV is a comparison of quartz from different suppliers. This comparison was made to determine if the electromigration could be related to bulk properties. In this case, each sample was given an equivalent surface finish and metallization. We took as a good indicator of crystal perfection and purity, the infrared absorption attributable to the presence of hydroxyl ions. For example, Dodd and Fraser have assigned numerous absorption bands in alpha-quartz to the presence OH associated with the alkali ions,¹⁵ and Brice and Cole reported that OH absorption correlates well with dislocation density.¹⁶ It is known that dislocations can provide channels for rapid diffusion in crystals. Therefore, it was not surprising that we found the transported charge correlated with the OH absorption coefficient.

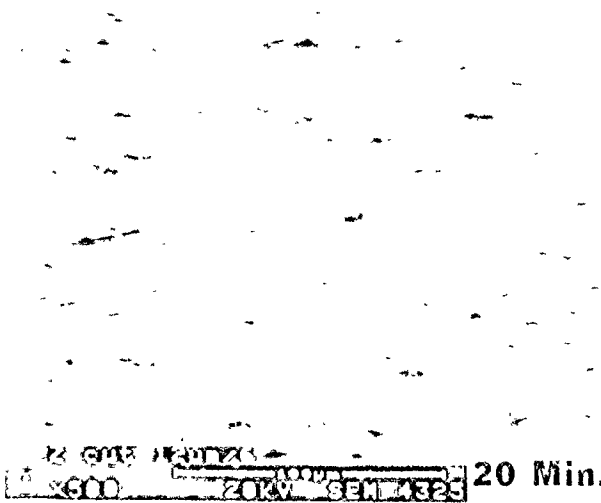


3 Min.



10 Min.

→ X



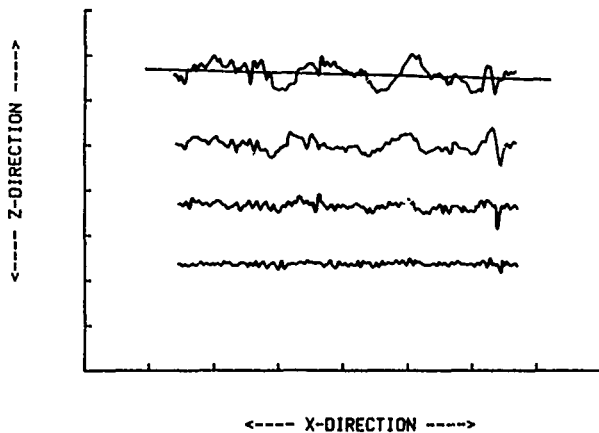
20 Min.



30 Min.

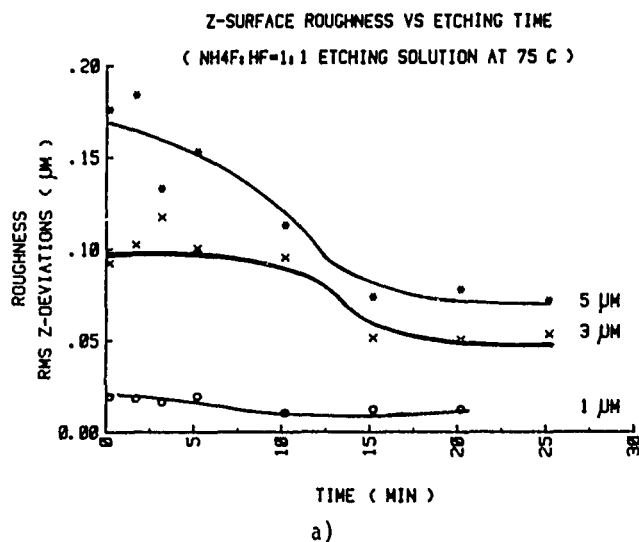
The figure shows the evolution of the surface morphology over time. At 3 minutes, the surface is covered with a dense, wavy pattern of small, rounded protrusions. By 10 minutes, the surface has developed a more pronounced, diamond-shaped grid pattern. At 20 minutes, the surface appears significantly smoother and less textured, with only faint, scattered protrusions visible. By 30 minutes, the surface has returned to a textured state, with a pattern of small, irregular protrusions. This sequence of images illustrates the dynamic nature of the surface morphology, which can change rapidly over time.

TYPICAL TALYSURF SCAN PROFILES

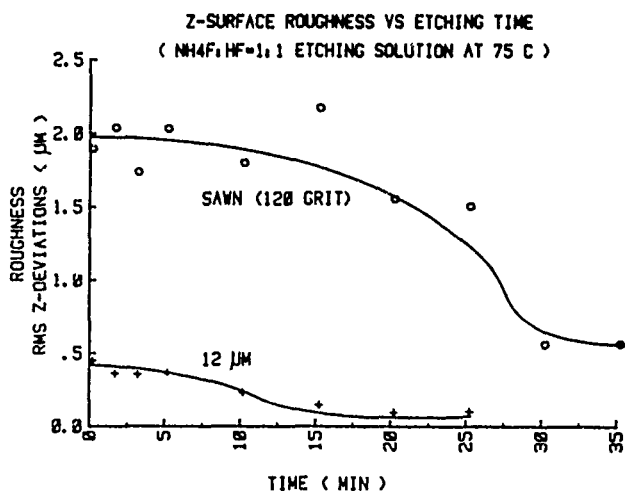


Typical as-recorded Talysurf scans. The root-mean-square of the Z-deviations from a best-fit straight line was equated to the roughness.

Figure 5.



a)



b)

Plots of surface roughness vs etching time for, a) 1 μm , 3 μm , and 5 μm surface treatments and b) 12 μm and sawn (120 grit) treatments.

Figure 6.

ACKNOWLEDGEMENTS

The authors thank William Washington for the finishing of samples, Ronald Brandmayr for the etching work, and Donald Boyce and Melvin Wade for the metal evaporations. Thanks are also due to John Vig and Joel Shappirio for helpful suggestions.

REFERENCES

1. J. G. Gualtieri, Proc. 39th AFCS, 247-254 (1985)
2. Richard K. Cook and Robert G. Breckenridge, Phys Rev 92 1419-1423 (1953)
3. J. Verhoogen, Am. Mineralogist 37, 637 (1952)
4. J. C. Brice Rev. Mod. Phys. 57 105-146 (1985)
5. R. B. Belser and W. H. Hicklin, Proc. 23rd AFCS, 132-142 (1969)
6. R. R. Sharp and E. L. Pace, J. Phys Chem Solids 31, 2275-2279 (1970)
7. SC-sticks were cut from zero degree R.H. Y-bars using orientation III see Roger W. Ward, Proc. 35th AFCS, 99-103 (1981)
8. J. C. King "Fundamental Studies of the Properties of Natural and Synthetic Quartz Crystals," (Final report, contract DA36-039, SC-64586, Prepared by Bell Telephone Laboratories, 10 June 1960)
9. Micro Abrasives Corp., 720 Southhampton Rd., Westfield, MA 01085
10. Buehler Ltd., 2120 Greenwood St., Evanston, Ill 60204
11. Zyp Coatings, P.O. Box 208, Oak Ridge, TN 37831
12. G. A. Wolff and J. D. Broder, Amer. Min. 45 1230-1242 (1960)
13. R. Brandmayr, USA LABCOM, Fort Monmouth, NJ 07703-5000, unpublished data
14. Dick Ang, Proc. 32nd AFCS, 282-285 (1978)
15. D. M. Dodd and D. B. Fraser, J. Phys. Chem. Solids 26, 673-686 (1965)
16. J. C. Brice and A. M. Cole, Proc. 32nd AFCS, 1-10 (1978)
17. J. Shappirio, J. Finnegan, R. Lux, D. Fox, J. Kwiatkowski, H. Kattelus, and M. Nicolet, J. Vac. Sci. Technol. A3 (6) 2255-2258 (1985)

IMPURITIES MIGRATION STUDY IN QUARTZ CRYSTAL RESONATORS BY USING ELECTROELASTIC EFFECT

R. Brendel*, J.J. Gagnepain, J.P. Aubry**

*Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté - Besançon
32 avenue de l'Observatoire - 25000 Besançon - France

** Compagnie d'Electronique et Piézoélectricité
44, avenue de la Glacière - 95100 Argenteuil - France

Summary

The electroelastic effect i.e. the frequency shift induced by application of an electric DC field on a quartz crystal operated at its resonance frequency can be used for the characterization of the ionic impurities contained in the crystal. The principles of the method, the experimental set-up and the procedure are described.

The frequency relaxation following the initial frequency shift appears to be dependent on different parameters, whose influence is studied. Among them the most important one is the temperature. The influence of the DC field strength is also examined. Comparisons are made between resonators made with different technologies such as B.V.A. (electrodeless), Q.A.S. (autosustained crystal with deposited electrodes) or regular resonators. The drastic difference between crystal swept and unswept shows that the electroelastic effect can be used as a test for the efficiency of the sweeping process.

An attempt of a theoretical analysis is presented in order to relate experimental observations to the characteristics of the diffusion process : thus it is shown that the amplitude of the relaxation curve is a function of the concentration of the impurities, the shape is related to their motion (mobility) and nature (migration energy).

Introduction

It is well known today that the performances of quartz crystal oscillators are limited, at least for the long term stability, by the quartz material itself. On the other hand, the sensitivity of the oscillator to various external perturbations such as accelerations, temperature fluctuations, irradiations ... limits its capabilities in severe environmental conditions as in military or space applications.

Large efforts have been achieved, for many years, in order to understand the sensitivity to irradiations, and, if possible, to reduce it. Many studies using different principles have shown that point defects contained in the crystal are responsible of some spurious effect observed on the quartz crystal oscillator behavior.

Among them high DC field relaxation method has the advantage of performing the measurements in the conditions of the regular use of the resonator in an oscillator.

Electroelastic effect

C. Hruska¹ pointed out in 1961 that a resonator operated at its resonant frequency and submitted to an electric field exhibits a frequency shift proportional to the strength of the DC field. This electroelastic effect also called polarizing effect is mainly due to third order nonlinear piezoelectric coefficients (electroelastic coefficients). Theory shows there exists 8 independent electroelastic coefficients. A phenomenological definition and numerical values obtained from the measurement of the electroelastic effect was given.^{2,3,4} Recently, starting from the works of Baumhauer and Tiersten^{5,6} and using experimental results published by Hruska and Kusters, expression and numerical values of the material nonlinear electroelastic coefficients was obtained.^{7,8}

In order to achieve more reliable results, a new set of measurements performed on a wider range of cuts are in progress and the corresponding calculated coefficients will be published. Preceding works^{7,8} have permit to show that the frequency shift $\Delta\omega$ of a plate resonator can be expressed as a function of the local electric DC field E

$$\frac{\Delta\omega}{\omega} = \frac{K}{2h\lambda} \int_{-h}^h E \cdot \cos^2\left(\frac{\omega x}{V}\right) dx \quad (1)$$

where x , extending from $-h$ to $+h$, is the abscissa along the normal to the plate of thickness $2h$. ω , V and λ are respectively the eigenfrequency, eigenvelocity and eigenvalue of the vibration depending upon the plate orientation (ϕ, θ) and vibration mode used. The coefficient $K(\phi, \theta)$ can be expressed in terms of linear and nonlinear elastic, piezoelectric and dielectric coefficients, among which the electroelastic coefficients are the most important ones.

The local DC electric potential $\Phi(x)$ varies from $\Phi(-h) = 0$ to $\Phi(h) = \Phi_0$ where Φ_0 is the applied DC voltage. In perfect crystals, $\Phi(x)$ varies linearly through the thickness of the plate so the electric DC field is constant

$$E_0 = -\frac{\partial\Phi}{\partial x} = -\frac{\Phi_0}{2h} \quad (2)$$

In such a case, equation (1) becomes

$$\frac{(\Delta\omega)}{\omega E_0 H} = \frac{K}{2\lambda} \quad (3)$$

Relation (3) expresses the homogeneous electroelastic effect i.e. the relative frequency shift per unit DC field.

The strength of the electroelastic effect calculated for the three modes of vibration and all the possible orientations of the plate, published earlier^{3,7}, shows that the singly rotated cut crystals do not present any electroelastic effect when the DC field vector is normal to the plate whereas this effect can reach magnitude of 10 pm/V (10^{-12} m/V) for doubly rotated cut. For example for C-mode one obtains -2.3 pm/V with SC-cut and -2.85 pm/V with FC-cut.

Relaxation process

Experimental measurement of the electroelastic effect in plate resonators is almost always followed by a slow quasiexponential decay of the frequency.^{4,8} The strong dependence with temperature and the order of magnitude of the decay time constant (several hours at ambient temperature) lead to the conclusion that this phenomenon is induced by the electromigration of ionic impurities contained in the crystal. The charge displacement modifies the local electric DC field which becomes non longer homogeneous through the plate but slowly changes with time until it reaches a steady state equilibrium. Nevertheless, equation (1) always holds, the DC field now depending both upon space x and time t . Hence, one can obtain the frequency as a function of time if the function $E(x,t)$ is known.

In order to build the diffusion model needed for the calculation of this function it must be recalled that the ionic conductivity of quartz is essentially due to the presence of interstitial cations M^+ such as Li^+ , Na^+ , H^+ ... trapped as compensating charges in the neighbourhood of $(Al^{3+})^-$ atoms substituted for Si^{4+} during the growth of the crystal and acting as negative charges.^{9,10} Thus, the crystal will be thought of as a medium in which positive charges (M^+) can move in a lattice where fixed negative charges $(Al^{3+})^-$ are evenly distributed. At the equilibrium the electrical neutrality needs there are as many cations M^+ as negative charges in the crystal. Furthermore it is assumed that the mobile charges can neither leave nor come into the crystal from the outside (blocking electrodes) so the net charge of the medium is always null as a whole.

Equations of diffusion

With the above assumptions, the electrodiffusion process is described by the following set of partial differential equations

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2} - \mu \frac{\partial (pE)}{\partial x} \quad (4)$$

$$\frac{\partial E}{\partial x} = \frac{p - n_0}{e} \quad (5)$$

$$E = - \frac{\partial \Phi}{\partial x} \quad (6)$$

where $p(x,t)$, $E(x,t)$ and $\Phi(x,t)$ are respectively the concentration of positive charges, the electric field and the electric potential.

D is the diffusion constant which is related to the mobility μ by Einstein's relation

$$D = \frac{kT}{e} \mu \quad (7)$$

k , T and e are the Boltzmann's constant, absolute temperature and electronic charge respectively. ϵ is the permittivity of the crystal along the direction of diffusion.

n_0 is the uniform concentration of negative charges so the conservation of the total charge inside the crystal is expressed by

$$\int_{-h}^h n_0 dx = \int_{-h}^h p \cdot dx = 2n_0 h \quad \text{for all time}$$

The applied potential Φ_0 gives the boundary conditions $\Phi(-h) = 0$, $\Phi(h) = \Phi_0$ for all time.

Equations (4-6) belong to nonlinear parabolic partial differential equations of "diffusion-convection" type for which there is no analytical solution known.

Nevertheless numerical solution of the stationary problem achieved in the past⁸ showed that the steady state frequency shift could be related to the migrating impurities concentration, furthermore, for large concentrations it was possible to obtain an analytical expression of that relation, namely

$$\frac{(\Delta\omega)_D}{(\Delta\omega)_H} = \frac{\pi^2 \epsilon E_0}{3he n_0} \quad \left(\text{if } \frac{n_0}{E_0} > \frac{e}{4eh} \right) \quad (8)$$

where subscripts D and H stand for steady state diffusion frequency shift and initial homogeneous electroelastic frequency shift respectively.

For the general case, in which concentration, potential and electric field all depend upon both space and time, even the numerical solution of the problem is difficult to obtain because this kind of equations leads often to numerical instabilities. Hence, the computation program presently used¹¹ can give consistent results only for weak concentrations such that

$$\frac{n_0}{E_0} < 10^{12} \text{ V}^{-1} \text{ m}^{-2} \quad (9)$$

However when condition (9) is verified, it is possible to calculate $E(x,t)$ which, in turn, allows one to obtain the frequency relaxation $\frac{\Delta\omega}{\omega}(t)$ by using relation (1). As a result, some not yet understood behavior of the high field frequency relaxation have been explained (see below experimental results and simulation of the relaxation process).

Experimental set-up

The quartz resonators under test are commercially available units of different technologies working at their resonant frequency in 5 or 10 MHz, third or fifth overtone oscillator.¹² The crystal is put in an oven whose temperature can vary from room temperature to about 100 °C. In order to avoid spurious effects due to temperature variations when the resonator is working far from its turnover point, the temperature is controlled by a home built thermostat with a resolution better than $\pm 5/10000$ °C. The high DC field is given by a voltage standard able to provide 1100 V with a stability of 10^{-5} /month. The voltage is applied alternately positive, null, negative and null (see Fig. 1) owing to a commutator driven by the computer. The period of the cycle depends on the time needed to

reach the steady state of diffusion. The frequency variations measured by a programmable counter are recorder by the computer.

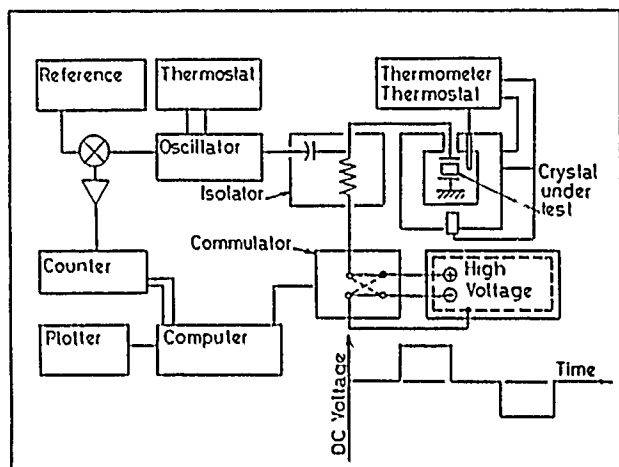


Fig. 1 - Experimental set-up

Experimental results and analysis of the steady state

Fig. 2 shows a typical record obtained with a BVA electrodeless SC-cut 10 MHz 3rd overtone resonator. The magnitude of the initial frequency shift leads to an homogeneous electroelastic effect of

$$\frac{(\Delta\omega)}{\omega E_0} \text{ measured} = 2.33 \text{ pm/V}$$

which is in good agreement with the one given by relation (3)

$$\frac{(\Delta\omega)}{\omega E_0} \text{ calculated} = 2.30 \text{ pm/V}$$

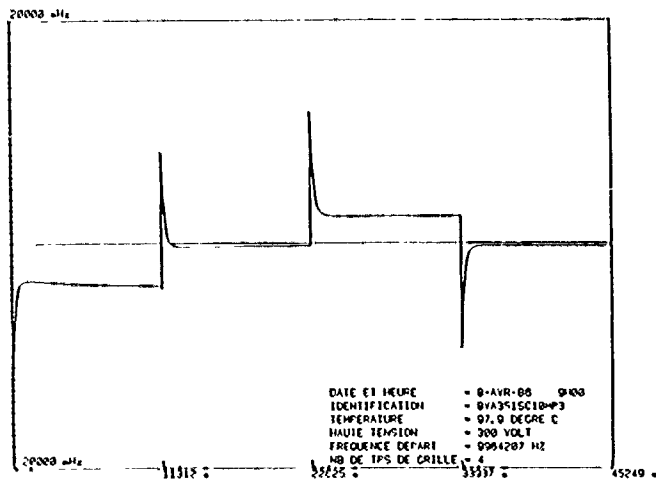


Fig. 2

Typical record of high field frequency relaxation

By using relation (8) an estimation of the impurity concentration can be obtained from the ratio between steady state and initial frequency shifts

$$\frac{(\Delta\omega)_D}{(\Delta\omega)_H} = 0.3 \text{ could correspond to } n_0 \approx 5.10^{18} \text{ charges/m}^3$$

it should be noted that this value corresponds to those impurities which are electrically charged and sufficiently small to move through the crystal, their concentration depends on their nature and the temperature according an Arrhenius' law

$$n_0 = N_0 \exp(-W_a/2kT) \quad (10)$$

where N_0 is the total impurities concentration for a given specie and W_a the corresponding activation energy.

Hence, if n_0 increases with temperature, the ratio $\frac{(\Delta\omega)_D}{(\Delta\omega)_H}$ given by relation (8) should decrease with it,

Figure 3 shows that if this assumption is verified for positive electric field it is not for negative field. This fact, not yet explained, shows that the behavior of the diffusion within the crystal is not symmetrical with respect to the direction of the electric field.

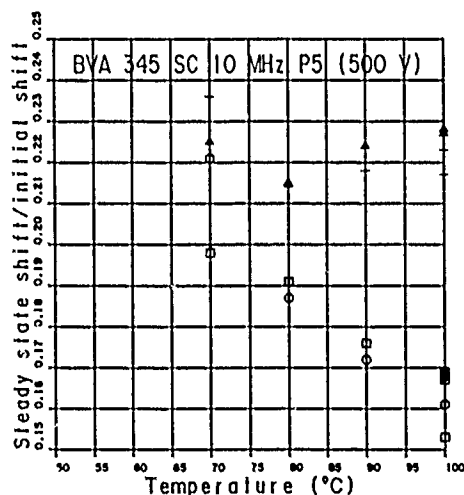


Fig. 3 - $(\Delta\omega)_D/(\Delta\omega)_H$ increases with temperature for positive electric field (squares and circles) but decreases for negative electric field (crosses and triangles)

This fact also occurs when studying $\frac{(\Delta\omega)_D}{(\Delta\omega)_H}$ as a function of the electric field strength, equation (8) plans that the ratio should increase with E_0 whereas Figure 4 shows that this law is only observed for negative electric field.

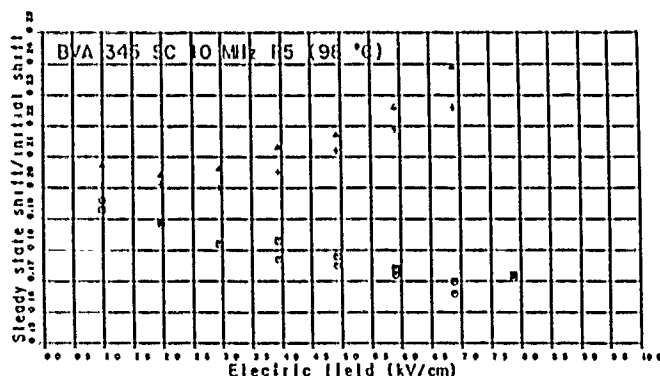
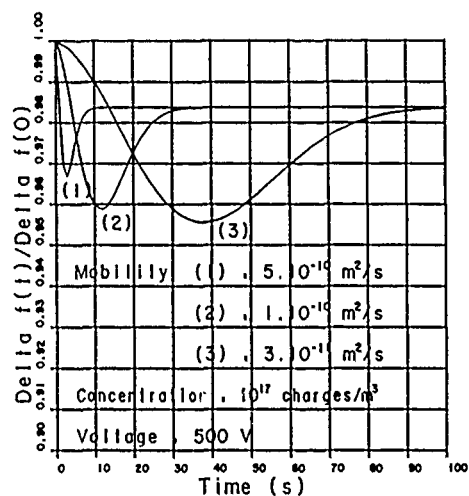
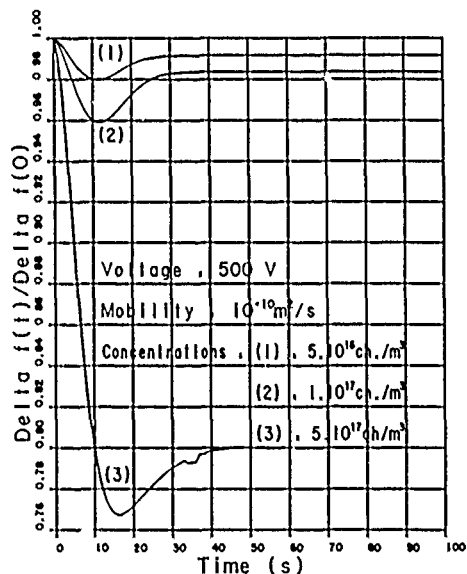


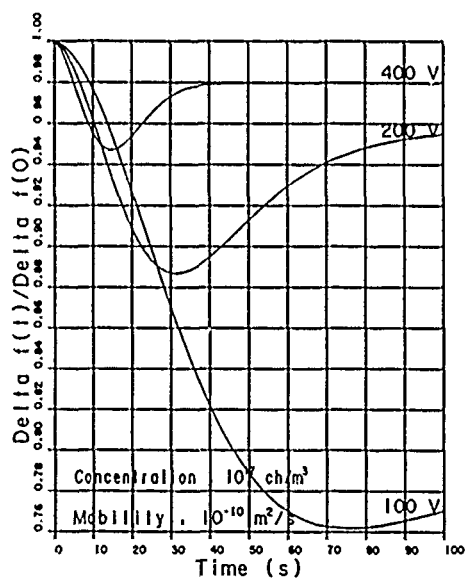
Fig. 4 - $\frac{(\Delta\omega)_D}{(\Delta\omega)_H}$ increases with electric field strength for negative values (crosses and triangles) but decreases for positive values (squares and circles)



(a)



(b)



(c)

Fig. 5

Numerical simulations of the frequency relaxation of a 1 mm thick plate by varying some parameters :

- a) mobility
- b) concentration
- c) DC voltage

Experimental results and simulation of the relaxation process

The early works^{4,8} on high field relaxation assumed that the frequency decay following the initial shift obeys an exponential law, nevertheless most of the records, as in figure 2, show a minimum of variation whose amplitude and position depend upon temperature and electric field strength.

These observations are confirmed by numerical simulation of the relaxation process represented on Figures 5.

They show that the position t_m and amplitude of the minimum depends upon the charge concentration, the mobility and the applied DC voltage. Note the strong dependence of the position t_m with the mobility, this fact can be verified experimentally because it is possible to increase the mobility by heating the crystal as in Figure 6 where the position t_m evolves as in Figure 5a when the temperature, therefore the mobility, increases.

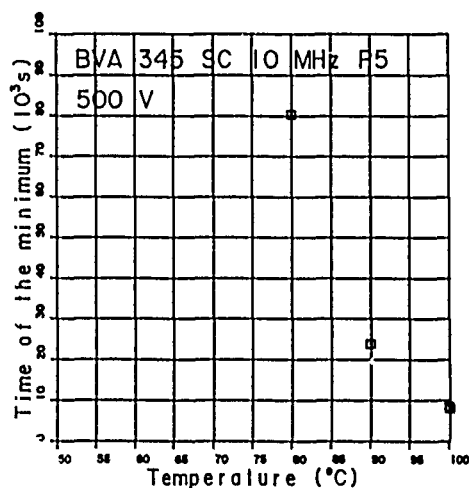


Fig. 6 - The position of the minimum of frequency variation strongly depends upon temperature

Furthermore, the points obtained by plotting $\ln(t_m)$ versus $1/T$ follow a straight line which can be identified with the Arrhenius' law

$$t_m = t_{m0} \exp (W_m/kT) \quad (11)$$

where W_m represents the migration energy, in the present case it is estimated to be 1.27 eV.

In some case, as in Figure 7, the frequency relaxation can present two minima (the experiment was too short to reach the second one).

This fact can be attributed to the presence of two kinds of impurities each of them moving with a different mobility. The simulation of such a condition represented in Figure 8 clearly shows the two minima observed experimentally.

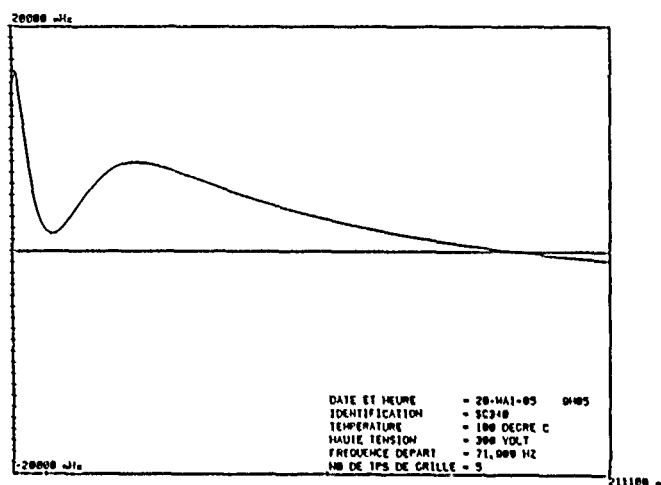


Fig. 7 - Experimental observation of frequency relaxation showing two minima ; they can be attributed to the presence of two kinds of migrating impurities

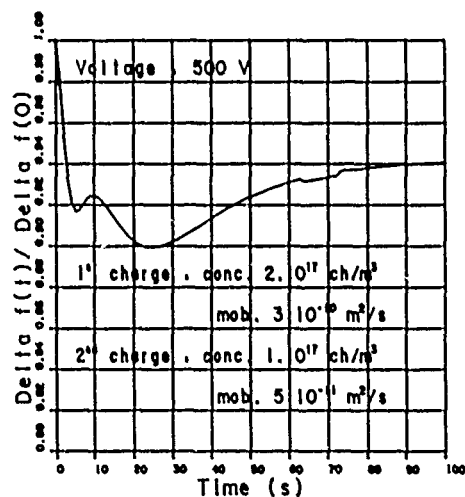


Fig. 8 - Numerical simulation of the frequency relaxation of a 1 mm thick plate containing two kinds of impurities

Other experimental features

Two interesting results of the high field relaxation are worth seeing.

One of them is dealing with the role of electrodes. In order to point out their responsibility in the frequency relaxation, two experiments have been performed : in the first one, the crystal mounted without adherent electrodes (BVA technology) behaves like in Figure 9a whereas in the second one the same crystal now directly electroded exhibits a thoroughly different response (see Figure 9b) as if the electrodes induced a supplementary much slower relaxation phenomenon whose mechanism is not yet entirely explained. It is obvious there that the steady state has not been reached at the opposite of Figure 9a and that the electrodes do not affect in any way the initial homogeneous electroelastic frequency shift.

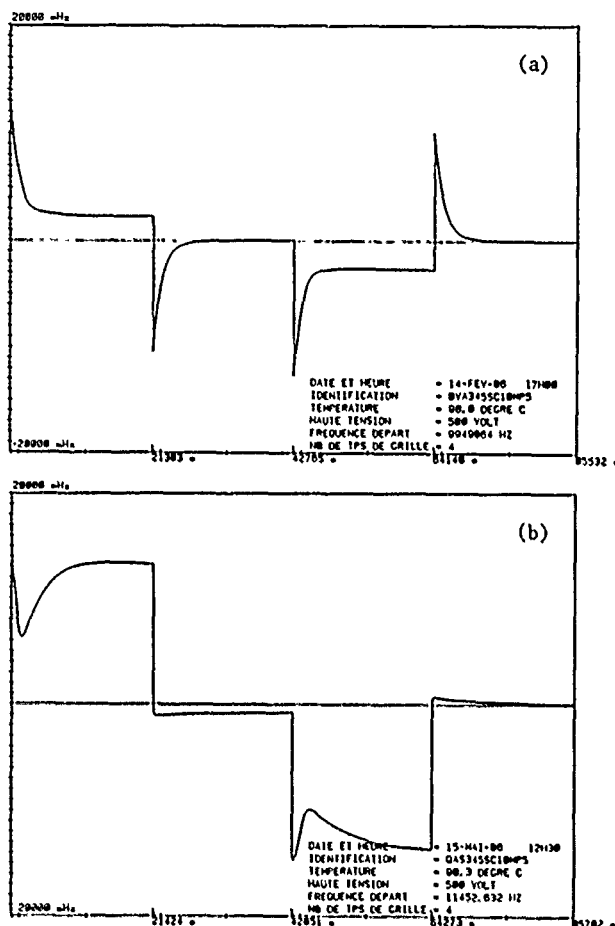


Fig. 9 - Frequency relaxations of the same resonator (a) without electrodes (b) with electrodes

Note that in the second case the steady state is not reached

The other notable feature concerns the influence of the sweeping process on the frequency relaxation. For this purpose two resonators of QAS technology taken in the same experimental conditions have been observed. The records of the Figure 10 show the drastic difference between unswept (a) and swept (b) material which does not present any relaxation as expected for an ideally pure crystal. Note that the sweeping neither modify the initial frequency shifts.

This observation shows that the high field relaxation measurement can be used as a method for estimating the efficiency of a sweeping process. In the above example numerical simulation indicates that, in such conditions, it remains less than 10^{16} moving charges/m³ (about a few parts for 10^{12} atoms Si). Furthermore, measurements made before and after the experiment show that the method do not modify in any way the electrical characteristics of the resonator (frequency and Q-factor).

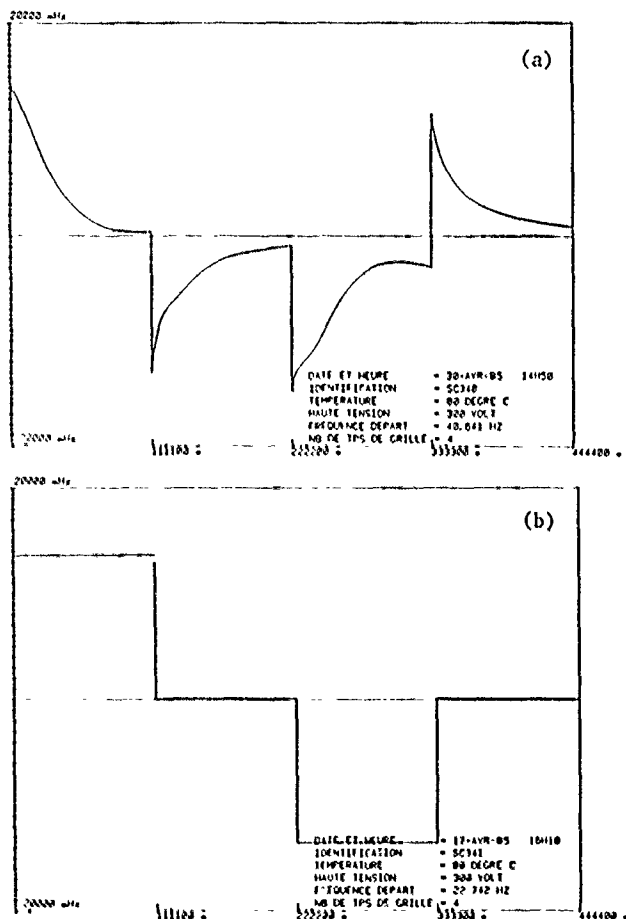


Fig. 10
Différence entre deux résonateurs de même technologie
(QAS 10 MHz troisième surtonne)
(a) matériau non balayé
(b) matériau balayé (procédé Sawyer)
Le dernier ne présente aucune relaxation telle qu'attendue
pour un matériau purement idéal

Conclusion

The high field frequency relaxation appears to be a very sensitive tool for the electromigration study in quartz crystal resonators. Simulations, when possible, show that concentrations as low as 10^{18} charges/m³ (a few ppb Si) can easily be measured by this method. Although many of the experimental behaviors have been explained with the help of numerical simulation some aspects remain to be clarified namely identification of the mobility or diffusion constant and migration energy (related to the kind of impurity) by numerical processing of the experimental data. This needs to improve the actual program to higher impurity concentration. The assumption made that charges do not enter nor leave the crystal is not yet confirmed by experimental results. This point could be settled by simultaneously recording frequency variations and DC current. Experimental set-up allowing such measurements is now under construction.

Acknowledgements

The authors wish to acknowledge the Compagnie d'Electronique et de Piezoelectricité (CEPE) which prepared and provided the resonators used in these experiments, and to thank G. Marianneau and C. Magnin for their help.

This work was made under contract 83/116 of the Direction des Recherches, Etudes et Techniques (DRET) - France.

References

- 1 C.K. HRUSKA, "The influence of an electric field on the frequency of piezoelectric cuts", Czechoslovak Journal of Physics, vol. 811, pp. 150-152, 1961.
- 2 C.K. HRUSKA, "The electroelastic tensor and other second order phenomena in quasilinear interpretation of the polarizing effect with thickness vibrations of α -quartz plates", Proc. of the 31st Annual Symposium on Frequency Control, pp. 159-170, 1977.
- 3 C.K. HRUSKA, A. KHOGLI, "Polarizing effect with doubly rotated α -quartz plates vibrating in thickness", IEEE Trans. on Sonics and Ultrasonics, vol. SU-25, n° 6, pp. 390-392, 1978.
- 4 J.A. KUSTERS, "The effect of static electric fields on the elastic constants of α -quartz", Proc. of the 24th Annual Symposium on Frequency Control, pp. 46-54, 1970.
- 5 J.C. BAUMHAUER, H.F. TIERSTEN, "Nonlinear electroelastic equations for small fields superposed on a bias", J. of the Acoust. Soc. of America, vol. 54, n° 4, pp. 1017-1034, 1973.
- 6 H.F. TIERSTEN, "Perturbation theory for linear electroelastic equations for small fields superposed on a bias", J. of the Acoust. Soc. of America, vol. 64, n° 3, pp. 832-837, 1978.
- 7 R. BRENDÉL, "Material nonlinear piezoelectric coefficients for quartz", J. of App. Phys., vol. 54, n° 9, pp. 5339-5346, 1983, and Erratum in J.A.P., vol. 55, n° 2, p. 608, 1984.
- 8 R. BRENDÉL, J.J. GAGNEPAIN, "Electroelastic effects and impurity relaxation in quartz resonators", Proc. of the 36th Annual Symposium on Frequency Control, pp. 97-107, 1982.
- 9 L.E. HALLIBURTON, M.E. MARKES, J.J. MARTIN, "Point defects in synthetic quartz: a survey of spectroscopic results with application to quality assurance", Proc. of the 34th Annual Symposium on Frequency Control, pp. 1-8, 1980.
- 10 J.C. BRICE, "Crystals for quartz resonators", Rev. of Modern Physics, vol. 57, n° 1, pp. 105-146, 1985.
- 11 J.P. BARDET, G. ORMANCEY, G. GODEFROY, "Simulation numérique unidimensionnelle de la dynamique d'injection de porteurs dans les diélectriques. Application au rutile", Journal de Physique, vol. 38, n° 3, pp. 345-352, 1977.
- 12 J.P. AUBRY, A. DEBAISIEUX, "Further results on 5 MHz and 10 MHz resonators with BVA and QAS designs", Proc. of the 38th Annual Symposium on Frequency Control, pp. 190-200, 1984.

EVALUATION OF MECHANISMS FOR LOW-DOSE FREQUENCY SHIFTS IN CRYSTAL OSCILLATORS

Terry M. Flanagan and Roland E. Leadon

JAYCOR

San Diego, CA 92138

and

Darwin L. Shannon

TRW

Redondo Beach, CA 90278

Abstract

Two possible causes for the observed frequency shift of a crystal oscillator due to a small radiation dose to the crystal have been analyzed. Removal of mass from the surface of the crystal by the radiation does not appear to be a plausible cause for the frequency shift. However, the electric fields and voltages created in the crystal by charge that is photoemitted from the walls of the can around the crystal and is then trapped in the crystal are relatively large and could possibly produce the observed frequency shift. These trapped charges produce both normal and tangential electric fields inside the crystal. Consequently, the difference in this radiation-induced frequency shift for SC-cut resonators, which are primarily sensitive to normal stresses, and for AT-cut resonators, which are primarily sensitive to tangential stresses, is less than the difference in their normal voltage sensitivities.

slopes observed at higher doses. These low-dose positive shifts are common to both glass- and metal-enclosed resonators (Refs. 5 and 6). Ionization-induced desorption of the surface contamination has been offered as an explanation. Two possible mechanisms may explain how a small dose can produce such a relatively large frequency error, namely, (1) ionization-induced removal of mass from the surface of the crystal which affects its resonant frequency, and (2) the creation of electric fields and voltages across the crystal due to charge emitted from and deposited in the crystal. The following analyses examine the plausibility of these mechanisms as the cause of the frequency shift.

In order to evaluate these mechanisms, the ionization energy and charge deposited on the crystal surface and in the bulk of the crystal from incident x-ray photons must be evaluated. Then the effect of the ionization and charge on the frequency shift in the crystal can be estimated.

2. Dose and Emission Parameters1. Introduction

The resonant frequency of high-quality quartz crystal resonators fabricated from swept synthetic quartz usually exhibits a negative shift under exposure to ionizing radiation for doses above a few kilorads (SiO_2) (Refs. 1-6). The shift is not exactly linear in dose but is often described by a slope of the order of 10^{-12} per rad (SiO_2) for well-aged resonators.

At doses below a kilorad (SiO_2), a positive frequency shift was noted by Flanagan and Wrobel (Ref. 2) which saturated at approximately one part in 10^8 in glass-enclosed resonators. Further investigations (Ref. 7) showed saturation of the positive frequency shift at approximately 300 rads for glass-enclosed resonators and no evidence of a positive shift for copper-enclosed resonators which had been subjected to high-temperature bakeout prior to sealing. This led these investigators to postulate that the lower-temperature processing and materials used in the glass-enclosed resonators led to surface contamination which was desorbed by ionization. The desorption would result in a decrease in the mass loading of the electrodes producing a positive frequency shift.

Subsequent data have shown that positive frequency shifts are observed for ionization doses as low as one rad (SiO_2) with rates of change per rad as much as several orders of magnitude larger than the negative

The parameters required for the subsequent analyses are: (1) the forward and backward electron emission currents from the walls of the can that surround the crystal and from the crystal itself, and (2) the photon-driven electron current through the bulk of the crystal, both calculated per unit dose in the crystal (SiO_2). These quantities are the result of incident x-ray photons which are energetic enough to penetrate the surrounding structure (oscillator cover, resonator oven, etc.) and the walls of the can that encloses the crystal. In this section, estimates are given for the emission parameters for use in the subsequent sections. The main structure of the can that enclosed the crystal is 0.025 cm of Cu with an inner coating of 2.5×10^{-5} to 1.2×10^{-4} cm of Ni followed by 10^{-4} cm of Au facing the crystal. The gold electrodes on the crystal, through which the frequency-controlling voltage is applied, are 7.5×10^{-6} cm of Au.

The photoemission properties of a surface are a function of the exterior material on the surface, provided the thickness of that material is more than the range in that material of the electrons excited by the photons. Otherwise, the emission properties are more characteristic of the underlying material. To estimate the emission properties of the relevant surfaces for this problem, the forward and reverse emission properties of Cu, Ni, Au, and SiO_2 were calculated using the QUICKE2 computer code (Ref. 8)

assuming that each material was more than one electron range thick. The incident photon spectrum was chosen to be a 15-keV blackbody and the intervening structure of the spacecraft and the oscillator up to the Cu can around the crystal was approximated by 1/4 inch of Al. Neither the incident spectrum nor the amount of intervening material is particularly crucial to this analysis because the emission currents will be normalized to one rad (SiO₂) deposited in the crystal. After the photons pass through any significant amount of material, the ratio of emission current to dose is fairly insensitive to the precise spectrum of the remaining photons. The important quantities are summarized in Table 1.

Table 1. Emission yields for one rad (SiO₂)
(15-keV blackbody through 0.25 inch of Al
and 0.01 inch of Cu)

Emitter	Cu	Ni	Au	SiO ₂
Forward yield from emitter (coul/cm ²)	3.0(-12) ^(b)	2.8(-12)	8.9(-12)	2.4(-13)
Average energy of electrons (keV) forward emitted from emitter material	37.1	38.2	39.4	36.3
Average range (cm) of forward emitted electrons in emitter material	4.7(-4)	4.7(-4)	2.6(-4)	1.25(-3)
Average range (cm) in SiO ₂ of electrons forward emitted from emitter material	1.6(-3)	1.6(-3)	1.9(-3)	1.25(-3)
Reverse yield from emitter (coul/cm ²)	1.85(-12)	1.7(-12)	6.9(-12)	1.0(-13)
Average energy of electrons (keV) reverse emitted from emitter material	32.5	33.1	37.2	31.6
Average range (cm) in SiO ₂ of reverse emitted electrons in emitter material	3.6(-4)	3.6(-4)	2.5(-4)	1.0(-3)
Average range (cm) in SiO ₂ of electrons reverse emitted from emitter material	1.2(-3)	1.2(-3)	1.8(-3)	1.0(-3)
Forward bulk current ^(a) density in SiO ₂ (coul/cm ²)	--	--	--	2.7(-13)
Reverse bulk current density in SiO ₂ (coul/cm ²)	--	--	--	1.5(-13)
Net bulk current density in SiO ₂ (coul/cm ²)	--	--	--	1.2(-13)

(a)The word "current" is sometimes used to denote time-integrated charge (coul).

(b)Numbers in parentheses indicate powers of 10.

As shown in Table 1, the emitted electrons have ranges in Ni and Au on the order of 4×10^{-4} cm and 2.5×10^{-4} cm respectively. Since this range in Au is considerably more than the thickness of the Au electrodes on the crystal (7.5×10^{-6} cm) the emission from the crystal, even in the region of the electrodes, can be taken as the emission from SiO₂, that is, 2.4×10^{-13} coul/cm² forward and 1×10^{-13} coul/cm² reverse, for one rad (SiO₂). Although the electron ranges in Ni and Au are somewhat larger than the thicknesses of Ni and Au on the inner wall of the can, the thicknesses are not negligible compared to the ranges. Therefore, the emission from the can walls will be a combination of the emission from Cu, Ni, and Au. For the subsequent analysis, it is assumed that the emission from the can inner walls is 50% due to Cu, 10% due to Ni, and 40% due to Au. Hence, the effective yields from the can walls are 5.3×10^{-12}

coul/cm² forward and 3.85×10^{-12} coul/cm² reverse for one rad (SiO₂). An approximate average range of these electrons in SiO₂ is about 1.7×10^{-3} cm.

3. Loss of Mass

The frequency of a precision crystal resonator is tuned to the desired frequency by vapor-depositing material on the crystal. From the observed change in frequency with deposited material, it is estimated that removal of only 10^{-3} of a monolayer of material by the one rad dose could account for the observed change of $\Delta f/f = 10^{-11}$. Therefore, the question is whether one rad (SiO₂) can remove 10^{-3} of a monolayer from the crystal. Assuming that the average distance between atoms in the crystal is $5 \text{ \AA} = 5 \times 10^{-8}$ cm, the surface density of atoms is 4×10^{14} atoms/cm². Therefore, removal of 10^{-3} /rad of a monolayer means removal of 4×10^{11} atoms/cm²/rad.

For this discussion, assume that the mechanism by which the dose removes the atoms is by ionizing the atoms, by capture of electrons that either are emitted from the inner surface of the enclosing can toward the crystal or are thermally excited in the crystal due to the deposited dose. Although the desorption probability is probably less than one per ionization, for these calculations a desorption probability of one is assumed.

For the first estimate, we will use the maximum emission from the can wall, that is, forward emission, 5.3×10^{-12} coul/cm²/rad, or 3.3×10^7 electrons/cm²/rad. Even if each of these electrons were captured by the first monolayer of the crystal and each captured electron caused an atom removal, the total removal (3.3×10^7 /cm²/rad) would be over four orders of magnitude too small (compared to 4×10^{11} /cm²/rad) to account for the observed frequency shift. Since many of these electrons will penetrate below the first monolayer, the number of these electrons captured in the first monolayer would be even smaller.

For another estimate, assume that the dose of one rad (SiO₂) ionizes the local atoms. Assuming an average energy of 5 eV to create one electron-ion pair in SiO₂, one rad (SiO₂) would produce an ionization density of about 2×10^{13} /cm³. [For comparison, the ionization rate in Si is 4.3×10^{13} /cm²-rad(Si)]. Thus, the number of ionization events in the first monolayer (5×10^{-8} cm) is only 10^6 /cm², which again is much smaller than the required density to explain the frequency shift.

In summary, it does not appear that mass removal due to ionization of the surface atoms is a plausible explanation for the observed frequency shift.

4. Radiation-Induced Electric Fields and Voltages

For the following discussion, consider the simplified geometry in Figure 1 for the crystal and the nearby interconnected (grounded) metal. The

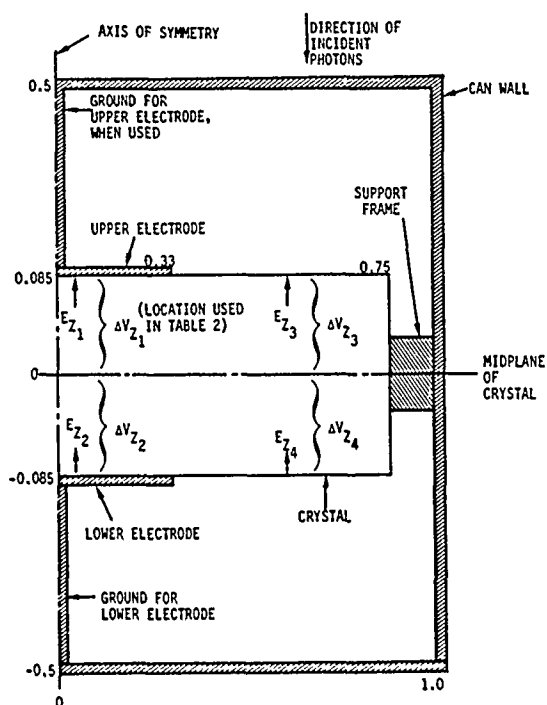


Figure 1. Assumed cylindrically symmetric geometry for analysis (dimensions are in centimeters; drawing not to scale).

crystal is assumed to be a short cylindrical section of SiO_2 located inside a concentric cylindrical can. The crystal is held in place by metallic supports on the periphery of the crystal. On the top and bottom surfaces of the crystal are two circular gold dots between which a voltage bias is applied for the normal voltage control of the crystal frequency. Calculations are presented here for the normal situation where the two electrodes have a constant voltage bias between them and one electrode is shorted to the can wall, and for a case where the upper electrode is floating and only the lower electrode is grounded to the can wall. For calculating the change in voltages due to the radiation, the voltage bias can be taken to be zero; that is, the electrodes are shorted together.

When photons penetrate the walls of the enclosing can, charge exchange occurs between the metal areas and the SiO_2 . This separation of charge causes electric fields and voltages across portions of the SiO_2 which will persist for as long as it takes the charge trapped in the SiO_2 to bleed off to some metal. If these radiation-induced voltages were exactly the same as applying a voltage between the two electrodes of the crystal, it would be straightforward to obtain a rigorous estimate of the resulting frequency shift. Unfortunately, this is not the case. For example, for the case when the electrodes are shorted together, the total voltage between the electrodes is obviously zero. However, the radiation-induced charges on the surface of the crystal and throughout its volume create local electric fields inside the SiO_2 which presumably will affect the

frequency of the crystal to some extent, even though their integral between the two electrodes is zero. Since there are no known data on frequency shifts of crystals for such nonuniform fields and voltages, the procedure that will be used to estimate the frequency shift is to calculate the voltage differences across specific volumes of the crystal and, from these, make an estimate of the possible frequency shift. For simplicity of analysis, and probably the largest predicted effect, it is assumed that the photons are incident on the system along the cylindrical axis of the crystal (see Figure 1). Hence, with the simplified geometry in Figure 1, the problem has cylindrical symmetry.

4.1 Charge-Separation Processes

The charge separation which occurs when the photons penetrate the can that surrounds the crystal can be associated with three separate processes: (1) forward and reverse emission from the can walls, (2) forward and reverse emission from the SiO_2 crystal, and (3) photon-driven electron currents through the bulk of the crystal. Since the gold electrodes on the crystal are much thinner than an electron range, all emission from them can be ignored.

4.1.1 Emission from Can Walls. As indicated in Section 2, the emission from the can walls will be a weighted emission from Cu, Ni, and Au, 5.3×10^{-12} coul/cm²/rad forward and 3.85×10^{-12} coul/cm²/rad reverse, for one rad (SiO_2). The electrons emitted from the can walls that strike the crystal will be imbedded in the SiO_2 over an electron range of $\approx 1.7 \times 10^{-3}$ cm (Section 2). Since the gold electrodes are so thin, this penetration depth applies both to the bare portion of the crystal and the electrode region. Secondary electron emission from the crystal is ignored in this analysis.

4.1.2 Emission from Crystal. From Table 1, the emission yields from SiO_2 for one rad (SiO_2) are 2.4×10^{-13} coul/cm²/rad forward and 1.0×10^{-13} coul/cm²/rad reverse. Since the gold electrodes are so thin, the emission from the electrode areas will be essentially the same as from the bare SiO_2 .

The electron emission from the SiO_2 comes from a thin layer on the crystal surface, an electron range thick ($\approx 1.7 \times 10^{-3}$ cm), and the corresponding positive charge density remains on the surface. However, that is not the net charge density on the surface, as discussed later.

4.1.3 Bulk Currents in Crystal. As the incident photons pass through the crystal, a net forward electron current is produced throughout the crystal. It is convenient to divide this net forward current into a forward bulk current and a reverse bulk current (Table 1). For one rad (SiO_2), these current densities on the front surface of the crystal are 2.7×10^{-13} coul/cm² net forward. (The word current will

often be used to indicate time-integrated charge motion.)

If there were no attenuation of the photon beam through the crystal, these bulk currents would be uniform throughout the crystal. As the excited electrons move forward and backward inside the crystal and are stopped in an electron range from where they were created, they leave behind positive charge layers an electron range thick just below the two surfaces of the crystal, a portion of the bulk currents are emitted to create the emission yield discussed in Section 4.1.2. The net surface charge density is the algebraic sum of the exterior incident current, the emitted current, and the forward and reverse bulk currents, that is, the net bulk current.

Since there is always some attenuation of the photon beam through the SiO_2 , the bulk currents near the far side of the crystal are smaller than the bulk currents on the front side of the crystal by the ratio $\exp(-L/\lambda)$ where L is the thickness of the crystal and λ is the attenuation length for the photons in SiO_2 . This attenuation produces essentially a uniform charge density in the SiO_2 equal to

$$\rho = J[1 - \exp(-L/\lambda)]/L \approx J/\lambda,$$

where J is the net bulk forward emission current at the front surface. For an estimated value of $\lambda = 2 \text{ cm}$ and $J = 1.2 \times 10^{-13} \text{ coul/cm}^2$ for one rad (SiO_2)

$$\rho = 6 \times 10^{-14} \text{ coul/cm}^3.$$

The net bulk current density at the rear surface of the crystal is the net bulk current density at the front surface $-\rho L$.

4.2 Charge Distributions in SiO_2

Figure 2 illustrates the currents and the net charge distributions in the SiO_2 . Since emission from the gold electrodes is assumed to be negligible, this distribution applies both to the bare SiO_2 and the electrode regions. The resulting electric fields and voltages are given in Section 4.3.

From Figure 2, the (positive) surface charge densities on the front (F) and rear (R) surfaces of the crystal are respectively

$$\begin{aligned} \sum_{F\text{SiO}_2} &= -J_{FWALL} + J_{R\text{SiO}_2} + J_{\text{NET BULK, FRONT}} \\ &= -5.3 \times 10^{-12} + 1.0 \times 10^{-13} + 1.2 \times 10^{-13} \\ &= -5.08 \times 10^{-12} \text{ coul/cm}^2 \end{aligned}$$

and

$$\begin{aligned} \sum_{R\text{SiO}_2} &= -J_{RWALL} + J_{F\text{SiO}_2} - J_{\text{NET BULK, FRONT}} + \rho L \\ &= -3.85 \times 10^{-12} + 2.4 \times 10^{-13} - 1.2 \times 10^{-13} \\ &\quad + \rho L \end{aligned}$$

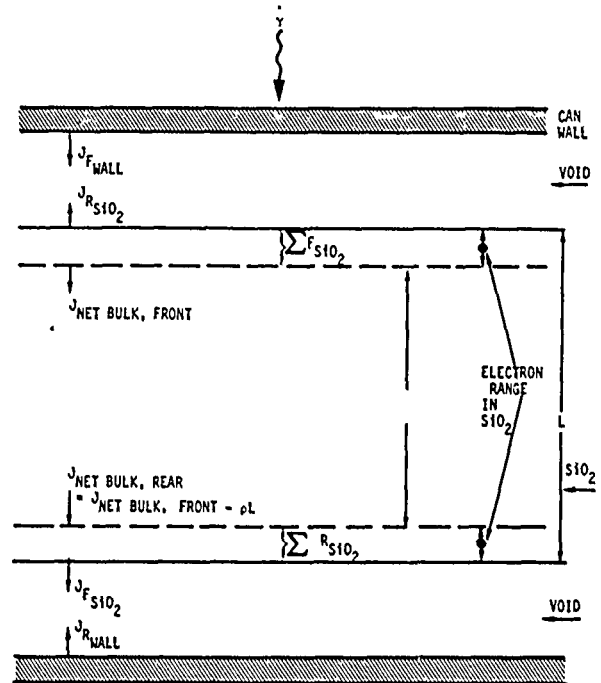


Figure 2. Illustration of currents and charges in crystal.

$$= -3.73 \times 10^{-12} \text{ coul/cm}^2 + \rho L.$$

For later calculation of fields and voltages (Section 4.3), it is convenient to keep ρL separated from $\sum_{R\text{SiO}_2}$. The image charge corresponding to

$\sum_{F\text{SiO}_2}$ and $\sum_{R\text{SiO}_2}$ is in the grounded metal of the system.

4.3 Electric Fields and Voltages

Two different methods are used for estimating the electric fields and voltages inside the crystal. First, in regions where the geometry and the charge distributions are close to planar, the fields and voltages are calculated by the standard planar formulas. These electric fields are obviously normal to the plane. Secondly, for the portion of the charge distribution that cannot be reasonably approximated by a planar geometry, a Poisson-solver computer code is used to solve for the electric fields inside the crystal. In general, these fields can be both normal and tangential to the surface of the crystal.

For calculating electric fields and voltages, it is convenient to separate the charge distribution given in Sections 4.1 and 4.2 into three components:

(1) the bulk charge distribution ρ and the compensating positive charge layer ρL on the bottom surface of the crystal, (2) the layers of charge beneath grounded electrodes on the crystal and the corresponding image charges in those electrodes, and (3) the remainder of the charges, that is, the charge layers on the bare surfaces of the SiO_2 crystal and under an ungrounded electrode, and the corresponding image charges in the grounded metallization of the system. The electric fields and voltages from the three components are directly additive since these effects are linear. The fields due to distributions (1) and (2) above are obtained by the planar analysis while those due to distribution (3) are obtained with the computer code. The calculations are done on a per rad basis.

For both the bulk charge and the charge layers beneath grounded electrodes, the fields and voltages can be approximated by parallel-plate analyses since the thicknesses of the charge layers are relatively small compared to their lateral dimensions.

The electric field near the bottom of the crystal due to the bulk charge ρ and its image charge ρL at the bottom of the crystal is

$$E_B(L) = \frac{\rho L}{K_{\text{SiO}_2} \epsilon}$$

where ϵ is the permittivity of free space and K_{SiO_2} is the relative dielectric constant for SiO_2 (≈ 3.9). For consistency, positive electric fields will always be directed upward in Figure 1. Also, voltage is defined as the negative integral of E . For $\rho = 6 \times 10^{-14}$ coul/cm³ (Section 4.1.4) and $L = 0.17$ cm (Figure 1)

$$E_B(L) = 0.30 \text{ V/cm}$$

This field decreases linearly to zero at the top surface. Therefore, the voltage across the crystal due to the bulk charge is

$$V = 1/2 E_B(L) L = -0.0026 \text{ V}$$

This field and voltage are independent of whether the upper electrode is grounded or floating.

The image charge density in a grounded electrode due to the charge in the SiO_2 just below the electrode is essentially equal to the charge density in the SiO_2 , that is,

$$\sum_{F_{\text{SiO}_2}} = -5.08 \times 10^{-12} \text{ coul/cm}^2/\text{rad},$$

for the top electrode when it is ungrounded and

$$\sum_{R_{\text{SiO}_2}} - \rho L = -3.73 \times 10^{-12} \text{ coul/cm}^2/\text{rad}$$

for the lower electrode since the charge sheet ρL has already been accounted for by the voltage due to the bulk charge, above.

The electronic field near the top electrode when it is grounded (positive upward in Figure 1) is

$$E_F = \sum_{F_{\text{SiO}_2}} / K_{\text{SiO}_2} \epsilon = -14.7 \text{ V/cm/rad}$$

and the field near the lower electrode is

$$E_R = - \sum_{R_{\text{SiO}_2}} - \rho L / K_{\text{SiO}_2} \epsilon = 10.8 \text{ V/cm/rad}.$$

These fields decrease approximately exponentially with distance into the SiO_2 away from the electrodes with a characteristic distance λ equal to an electron range in SiO_2 , ($\approx 1.7 \times 10^{-3}$ cm). Therefore, the voltage across the layer near the upper grounded electrode is

$$V_F = 0.025 \text{ V/rad}$$

and across the layer near the lower electrode is

$$V_R = -0.018 \text{ V/rad}.$$

Since the distance from the bare SiO_2 on the crystal to the can walls is comparable to or greater than the distance from the bare SiO_2 to other metal in the system, for example, the electrodes and the edge supports for the crystal, a planar analysis of the voltages due to the charge trapped in the bare SiO_2 , and under the upper electrode when it is floating, would be inaccurate. Consequently, the fields and voltages resulting from these charges were calculated using a computer code PRECHG. This code solves Poisson's equation on a cylindrically symmetric finite-difference r - z grid for fixed charges in a dielectric inside a cylindrical conducting can with any other cylindrically symmetric metal in the volume either grounded to the can or electrically floating. Calculations were made with the charges

$$\sum_{F_{\text{SiO}_2}} \text{ and } \sum_{R_{\text{SiO}_2}} - \rho L$$

on the upper and lower surfaces of the crystal simultaneously for two cases: (1) with the upper and lower electrodes grounded to the can wall, and (2) with the upper electrode floating.

After the code reaches a satisfactorily converged solution, it prints out the electric fields in both the z and r directions at all the mesh zones in the simulation volume. Thus, one could calculate the voltages between any two points in the volume. However, as discussed in Section 4.0, the electric fields that result from these charge distributions do not have the same spatial distribution as the fields that result from a voltage bias between the crystal electrodes. Consequently, it is not obvious which fields or voltages in the crystal would be most representative for trying to estimate the resulting

frequency changes. Therefore, to cover a range of possibly sensitive regions, the following quantities have been selected for comparison (see Figure 1).

1. Electric fields along the axis of symmetry (z direction) in SiO_2 adjacent to the upper and lower electrodes.
2. The voltages from the midplane of the crystal to the upper and lower electrodes.
3. Electric fields in z direction just inside the upper and lower surfaces of the SiO_2 at a radial location midway between the edge of the electrode and the outer radius of the crystal.
4. The voltages from the midplane of the crystal to its upper and lower surfaces at the same radial position as in (3).
5. Profiles of tangential electric fields at three radial locations in the crystal.

The values for items (1) to (4) are given in Table 2 for both electrodes grounded and for the upper electrode floating. Tangential fields when both electrodes are grounded and for the upper electrode floating are shown in Figure 3.

Table 2. Calculated normal electric fields and voltages in crystal due to charge on surface of crystal.

	Both electrodes grounded	Upper electrode floating
E_z (V/cm) ^(a) (per rad)		
(1) On centerline, near upper electrode	-1.4	+18.0
(2) On centerline, near lower electrode	+1.3	+20.0
(3) $r = 0.5$ cm, near top surface	+10.1	+10.0
(4) $r = 0.5$ cm, near bottom surface	-6.5	-6.0
ΔV_z (V) ^(b) (per rad)		
(1) Near centerline, crystal midplane to upper electrode	+0.073	-1.55
(2) Near centerline, lower electrode to crystal midplane	-0.073	-1.66
(3) $r = 0.5$ cm, crystal midplane to upper surface	-0.51	-0.54
(4) $r = 0.5$ cm, lower surface to crystal midplane	+0.21	+0.16

(a) Positive E_z is upward in Figure 1.

(b) $\Delta V_z = - \int_{z_1}^{z_u} E_z dz$ where z_u is above z_1 .

4.4 Frequency Shift Due to Radiation-Induced Voltages

The voltage sensitivity for SC-cut resonators has been measured to be 7×10^{-9} V while for AT-cut crystal the coefficient is about 1.1×10^{-10} (Ref. 9). This frequency shift is for a voltage applied between the two electrodes on the crystal, that is, the only non-neutral charge regions in the system are the upper and lower electrodes. Moreover, the resulting electric fields in the oxide are confined mainly to the region between the electrodes, and the axial component of the field is monopolar and more or less constant between the electrodes. As indicated

previously, the electric fields from the radiation-induced charge distributions estimated in Section 4.3 have none of the above characteristics. Charges and electric fields are distributed nonuniformly throughout the crystal and, for both electrodes grounded, the electric field between the electrodes is necessarily bipolar to give a zero voltage between the

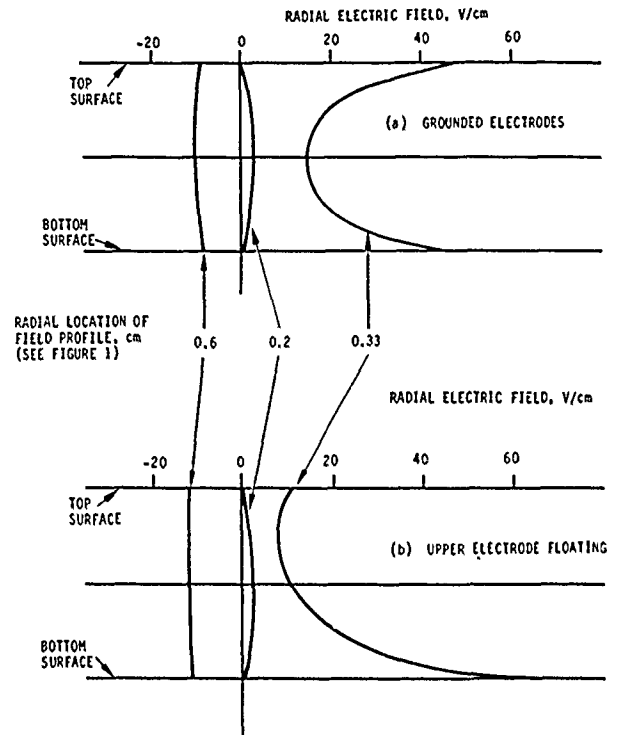


Figure 3. Profiles of radial electric field inside crystal.

electrodes. Consequently, for simplicity in estimating these fields and explaining the observed radiation-induced frequency shifts, it will be assumed that any axial voltage across any portion of the crystal will produce the same frequency shift as an applied bias. In other words, a large axial electric field across a small thickness of the crystal is assumed to produce the same frequency shift per volt as a small field across a larger thickness. With this assumption, the observed frequency shift $\Delta f/f = 10^{-11}$ could be explained by voltages of 1.4×10^{-3} V for SC cuts and ~ 0.1 V for AT cuts.

From Section 4.3, the voltage across the crystal due to the bulk charge ρ is -0.0026 V/rad while the voltages across the thin charge layers adjacent to grounded electrodes were 0.025 V/rad for a grounded upper electrode and -0.018 V/rad for the lower electrode. Moreover, for one rad from Table 2, the voltages from the crystal midplane to its upper and lower surfaces due to the charge layers on the surface of the crystal are on the order of 0.07 V to 0.5 V for both electrodes grounded and up to 1.6 V for the upper electrode floating. Although some of these voltages have opposite signs and thus their effects on the frequency may partially cancel each other, the upper ranges of these voltages are so much larger than the control voltages ($\approx 10^{-1}$ V to 10^{-3} V) required to

produce a frequency shift of $10^{-11} \Delta f/f$ that it is quite plausible that the asymmetries in the problem will produce a net frequency shift of the magnitude observed. However, one difficulty with this explanation is that it would imply that SC-cut resonators should have much larger radiation-induced frequency shifts than AT-cut resonators, assuming that their geometries are similar, whereas it has been observed that the shifts are comparable in both types of resonators. This effect could be explained by considering the tangential electric fields (Figure 3) as well as the normal fields (Table 2). Since the tangential and normal electric fields are comparable in magnitude, the tangential fields could be the primary cause of the observed shifts in the AT-cut resonators while the normal fields are responsible for the shifts in the SC-cut resonators.

As an illustration of the importance of the tangential electric fields, Eer Nisse (Ref. 10) has measured a frequency shift due to a tangential stress (S) of

$$\frac{\Delta f}{f} = 2.75 \times 10^{-12} S$$

where S is in dyne/cm². From Ref. 11,

$$S = -e_{11}E$$

where $e_{11} = 0.173 \text{ C/m}^2$ is the stress element for quartz. Using a rather small value of tangential $E = 2 \text{ V/cm} = 200 \text{ V/m/rad}$ from Figure 3, $S = -34.6 \text{ J/m}^3 = -346 \text{ dyne/cm}^2$. The corresponding frequency shift is about $10^{-9}/\text{rad}$, that is, much larger than the observed shift of the order of $10^{-11}/\text{rad}$.

4.5 Discussion

The analyses presented here have calculated the expected mass-desorption caused by ionization of surface contamination layers based on accepted (and experimentally tested) codes for gamma and x-ray kinetic charge transfer and on the assumption of one desorption for each surface ionization event. The calculation leads to the conclusion that a desorption coefficient of the order of 10^4 molecules per surface ionization is required to explain the data on the basis of mass desorption alone.

Using the same calculations for kinetic energy transfer, and ranges of charged particles, the electric fields in the crystal have been estimated. From relationships between the electric field and the frequency, the conclusion is reached that the charge trapped in the crystal is the most likely cause of the observed frequency shifts.

It is to be noted that both Norton (Ref. 6) and Ho (Ref. 5) have reported that the low-dose frequency shifts are process dependent, and larger shifts are observed from processes in which higher levels of contamination are expected. It is likely that the presence of surface contamination loss affect the amount of trapped charge. It is also possible that, if the contamination layer becomes sufficiently thick,

mass desorption may play a significant role in low-dose frequency shifts.

The analysis presented here implies that, at least for properly cleaned surfaces, efforts to control or reduce low-dose frequency shifts should focus on controlling trapped charge.

References

1. R. A. Poll and S. L. Ridgeway, IEEE Trans. Nuc. Sci., NS-15, Dec. 1966, pg. 10.
2. T. M. Flanagan and T. F. Wrobel, IEEE Trans. Nuc. Sci., NS-16, Dec. 1969, pg. 130.
3. E. Euler, et al., Proc. 32nd Annual Symposium Freq. Control, 1978.
4. T. J. Young, O. R. Koehler, and R. A. Adams, Proc. 32nd Annual Symposium Freq. Control, 1978.
5. J. Ho, FEI Test Reports, private communication.
6. J. Norton, et al., IEEE Trans. Nuc. Sci., NS-31, Dec. 1984, pg. 1230.
7. T. M. Flanagan, IEEE Trans. Nuc. Sci., NS-21, Dec. 1974, pg. 390.
8. S. H. Rogers and A. J. Woods, "Multiple-plate Modification of the QUICKE2 Electron Emission Code," DNA Report 4064T, June 1976.
9. R. E. Feller and J. R. Vig, "Fundamental Mode SC-cut Resonators," Proc. 34th Annual Freq. Control Symposium, May 1980, pg. 187.
10. E. P. EerNisse, Proc. 29th Annual Freq. Control Symposium, 1975, pg. 1.
11. V. E. Bottom, Introduction to Quartz Crystal Unit Design, Van Nostrand Reinhold Company, New York, NY, 1982.

Low and Medium Dose Radiation Sensitivity of Quartz Crystal Resonators with Different AL-Impurity Content

J. J. Suter and R. H. Maurer

The Johns Hopkins University
Applied Physics Laboratory
Laurel, MD 20707

In this paper we will discuss the results from low (4.0 Rads) and Medium (1000 Rads) radiation tests on AT Quartz Crystal resonators with AL impurity content of 0.79 and 8.3ppm. The radiation tests were conducted with two Cobalt 60 sources which generate photons with a mean energy of 1.25 MeV. The dose rate at the quartz crystal blanks was approximately 4.0 Rads (Si) per hour, which represents approximately a factor of 10 acceleration on the dose rate experienced in the radiation environment of a low earth orbit. A mathematical analysis of the frequency susceptibility versus total accumulated dose revealed that a least squares fit of the df/f versus accumulated dose (R) data is of the type:

$$(df/f(R)) = AR^B,$$

where A and B are real rational coefficients. Some modest interpretation of this result will be presented, together with data on the Allan variances, and aging rates of the oscillators during and after the radiation tests.

I. Introduction

The Johns Hopkins University Applied Physics Laboratory has been involved in studies relating to the radiation sensitivities of Quartz Crystal Resonators for several years (references 1 and 2). Lately, these studies have concentrated on the frequency susceptibility of AT Quartz Crystal Resonators to low earth orbit radiation levels. Typically, the total accumulated radiation doses at the quartz crystal blanks for oscillators mounted in spacecraft in a low earth orbit, do not exceed 1000 Rads(Si) per year. Although low earth orbit radiation includes contributions by proton, electron and Bremsstrahlung, it has been shown that the frequency sensitivity (df/f) of AT Quartz Crystal Resonators can be accurately modeled with gamma radiation from Cobalt 60 sources (reference 1).

Reference 2 (reported at the 1985 PTTI Conference) showed that the rate of change of the frequency shift with radiation dose, df/f per rad, is a function of the total accumulated dose and that this slope (df/f per rad) is the largest for low [4.0 Rads(Si)] accumulated doses. Reference 2 reported data on only one AT cut quartz crystal resonator over an interval from 0 to 1000 Rads(Si) at a dose rate of approximately 4 Rads(Si) per hour. We have repeated this experiment on two additional AT cut resonators, one with an aluminum impurity content which is the same as the first (0.79 ppm) and one with an aluminum impurity content an order of magnitude greater (8.3 ppm). Since the responses of the three resonators appeared to be similar during the 1000 Rads(Si) exposure and, to a lesser extent, during recovery following the end of irradiation, we judged that it would be worthwhile to perform a least squares fit on the df/f versus accumulated dose data. Both the data from the three experiments and the mathematical analysis of the results are presented in this study. Besides the results on df/f per rad, data on

drift rates and Allan variances during and after irradiation are presented.

Our efforts in these studies are directed at extending the work reported previously in References 3-6. In general, we recognize that the frequency susceptibility of AT quartz crystal resonators cannot be extrapolated from high dose [≥ 10 kRads(Si)] results. We accept the fact that for best performance in a radiation environment, one should start with swept, synthetic quartz. We are interested in investigating the effects of aluminum impurity content, resonator processing and quartz crystal resonator technology on the performance of resonators, particularly in the low earth orbit (LEO) radiation environment. To date our experiments have shown that while quartz resonators are sensitive to radiation, they are also robust, continuing to perform after several exposures to severe radiation environments.

II. Experimental Procedures

The three test bed oscillators used in the experiments were all equipped with AT cut quartz crystal resonators. The resonators were previously used in radiation susceptibility tests in which they were subjected to high doses of gamma ray (up to 20 KRads), and proton (15 KRads accumulated doses) irradiation (references 1 and 2). As is shown in Table I, each of the test bed oscillators was characterized by measuring its aging and 10 second Allan variance. We found that these data were reproducible for each of the oscillators as compared with similar data taken in previous radiation experiments (References 1 and 2). The test configuration shown in Figure 1 was used for the short and medium term radiation experiments. Photons with a mean energy of 1.25 MeV from two Cobalt 60 sources were incident on the Quartz Crystal Oscillators. The thickness of the shielding between the sources and the quartz crystal blank was calculated to be about 9.0 gm/cm² of equivalent aluminum shielding.

As is shown in Figure 2, the 5.0 MHz RF output from the AT quartz test beds was applied to an APL developed mixer in which it was combined with an RF signal from a laboratory quartz crystal oscillator standard. The mixer, in turn, generated a 400 Hz beat frequency which was applied to the HP data acquisition system. By using an uninterruptable power supply we were able to conduct the data acquisition over long periods of time without suffering power surges or interruption. The computer software developed at APL has the capability of monitoring the frequency shift (df/f) as a function of time and/or accumulated radiation dose, and 1, 10, or 100 second Allan variances. In the experiments described in this paper, data on the oscillator frequency shift (df/f) were collected in 10 and 100 second intervals. Furthermore, the oscillator frequency shifts were also monitored by means of an analog strip chart recorder as a function of real clock time.

Table I
Pre-radiation at resonator characteristics.

AT RESONATOR* ID	QUARTZ BAR	AUTO CLAVE	ALUMINUM CONTENT (ppm)	24 HOUR AGING RATE†	10 SEC ALLAN VARIANCE†
33820	K13F	A6 26	0.79	2.17×10^{-11}	$< 4.0 \times 10^{-12}$
34632	K11K	H39 24	83	2.05×10^{-10}	$< 6.0 \times 10^{-12}$
34744	K13F	A6 26	0.79	2.24×10^{-10}	$< 1.5 \times 10^{-12}$

* PREMIUM - Q. SWEPT, CULTURED, SYNTHETIC QUARTZ PRODUCED BY SAWYER RESEARCH PRODUCTS.
RESONATORS WERE MANUFACTURED BY BLILEY ELECTRIC COMPANY, RESONATORS WERE PRE-CONDITIONED
TO 20 KRADS (SI).
† MEASURED IN JUNE 1985

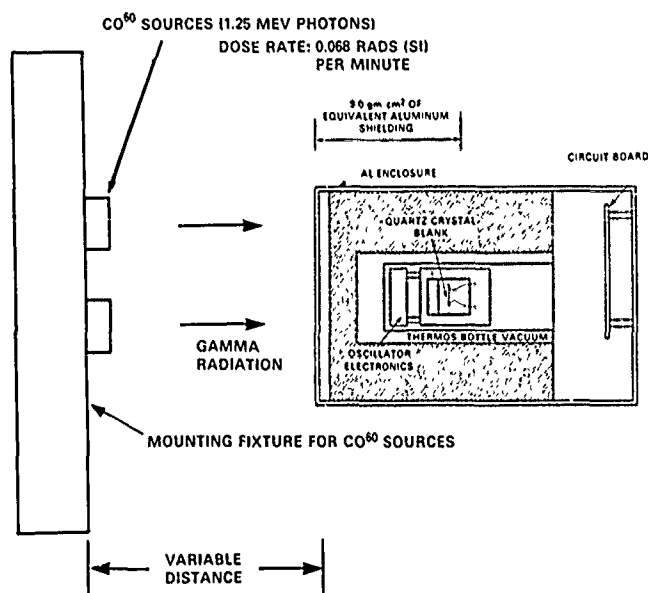


Fig. 1 Quartz crystal oscillator radiation test configuration.

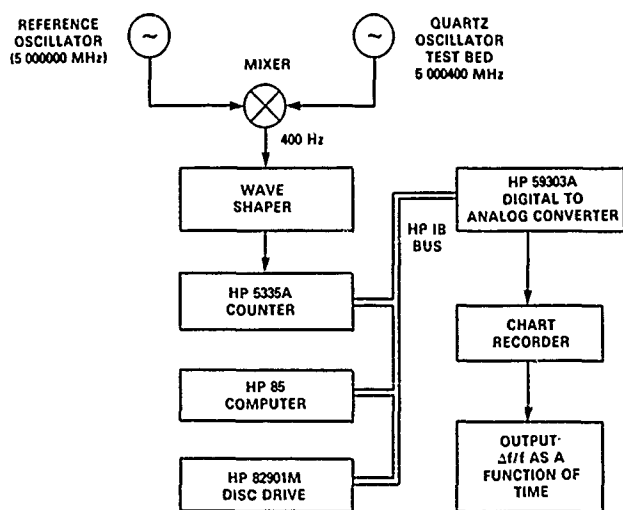


Fig. 2 Data acquisition system for oscillator radiation test.

As reported previously, the aluminum content of each resonator was determined from a sample of the quartz bars by Oklahoma State University (Reference 2 and 7).

III. Low Dose Gamma Ray Tests

Table 2 shows the results of the low dose [0-4 Rads(Si)] experiments carried out at a dose rate of 4.1 Rads(Si) per hour during June 1985 on each of three AT resonators included in this study. The results of these low dose experiments were discussed more extensively in Reference 2.

Succinctly, the df/f per rad response for each resonator is shown in the top row, columns 3, 4 and 5 of Table 2. Even though resonators 33820 and 34744 have the same aluminum content (they were cut from the same bar), they had quite different responses to the 4 Rads(Si) exposure. Sample number 33820 had a positive shift of 4.92×10^{-11} per rad while resonator number 34744 had a negative shift of -4.45×10^{-11} per rad following a positive excursion of $df/f = +2.43 \times 10^{-10}$. Resonator 34632, which has an aluminum impurity content an order of magnitude greater than the other two resonators, had a positive shift of 3.22×10^{-11} per rad. This type of confusing data - resonators of lower aluminum content having greater radiation sensitivity, positive and negative df/f values or both during the same exposure - is representative of the results experienced several years ago at APL. Reference 1 summarized the results and found that a correlation existed between radiation sensitivity and the processing lot of the AT resonators. We hypothesize that resonator processing methods are important at low doses when resonator surface effects are not completely dominated by resonator bulk effects.

IV. Medium Dose Gamma Ray Tests

The same test configuration shown in Figure 1 for the low dose tests was employed during the later part of 1985. Cobalt 60 exposure of the AT resonators and data acquisition took place continuously at 3.8-3.9 Rads(Si) per hour dose rate for approximately 260 hours. The results of the experiments on resonators 33820, 34632 and 34744 are summarized in Tables 2 and 3 as well as Figures 3, 4 and 5.

Table 2 shows that the df/f per Rads(Si) slope decreases by one to two orders of magnitude as the dose range increases from a few rads to 500-1000 Rads(Si). Figures 3, 4 and 5 show this effect graphically. For each of the three resonators, the df/f response to the gamma rays saturates as the dose increases. The least square fit slopes for the 0-40 Rads(Si) and 500-1000 Rads(Si) ranges (also listed in Table 2) are shown on the figures. From the first 4 Rads(Si) to the last 500 Rads(Si), the slopes decrease by factors of 16 (34632) to 86 (33820). The least squares functional fits to the df/f versus dose data will be discussed in Section VI.

The net shift at the end of the 1000 Rads(Si) exposure is:

Table II
 $\Delta f/f$ per rad (Si) for three at resonators.

AT RESONATOR SERIAL NUMBER	RESONATOR AL - CONTENT [ppm]	TOTAL ACCUMULATED DOSE RANGE [RADS (Si)]	DOSE RATE [RADS (Si)/hr]	$\Delta f/f$ PER RAD OVER RANGE
33820	0.79	0 - 4	4.1	4.92×10^{-11}
		0 - 40	3.9	3.31×10^{-11}
		500 - 1000	3.9	5.72×10^{-13}
34632	8.3	0 - 4	4.1	3.22×10^{-11}
		0 - 40	3.84	3.10×10^{-11}
		500 - 1000	3.84	2.04×10^{-12}
34744	0.79	0 - 4	4.1	-4.45×10^{-11} (*)
		0 - 40	3.80	1.07×10^{-11}
		500 - 1000	3.80	8.49×10^{-13}

(*) FOLLOWING POSITIVE EXCURSION OF $+2.43 \times 10^{-10}$

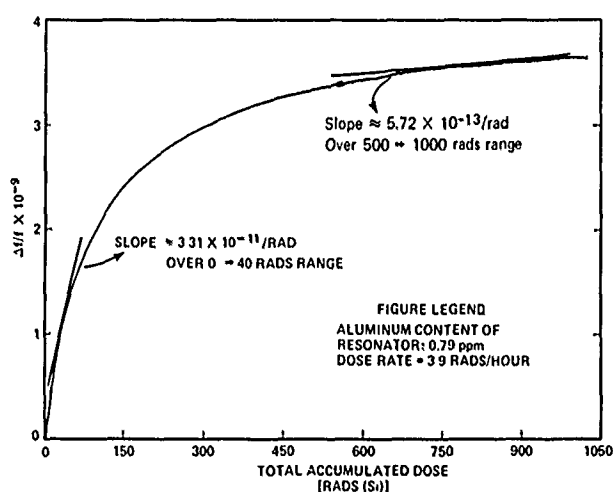


Fig. 3 Long-term low-dose rate radiation test on 33820 resonator.

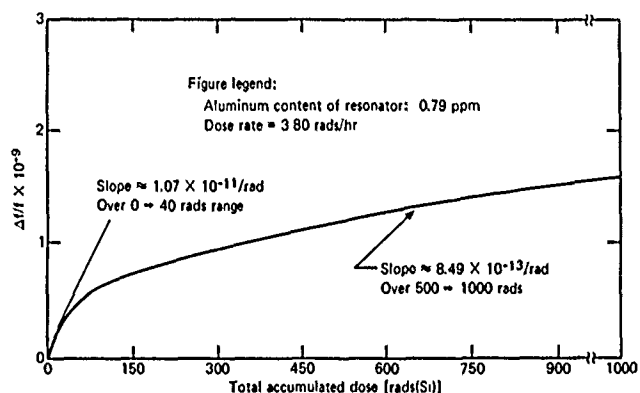


Fig. 5 Long-term low-dose rate radiation test on 34744 resonator.

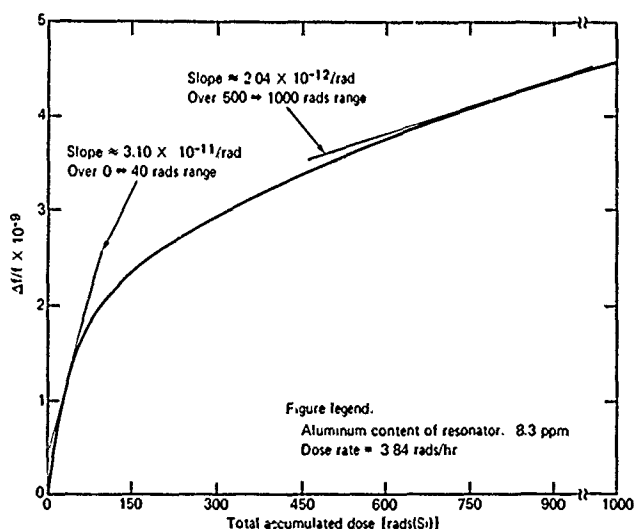


Fig. 4 Long-term low-dose rate radiation test on 34632 resonator.

$$\begin{aligned} df/f &= 3.5 \times 10^{-9} \text{ for 33820 (0.79ppm Al)} \\ df/f &= 4.5 \times 10^{-9} \text{ for 34632 (8.3ppm Al)} \\ df/f &= 1.6 \times 10^{-9} \text{ for 34744 (0.79ppm Al)} \end{aligned}$$

We now see that resonator 34632, which has the highest aluminum impurity content, also has the greatest net df/f shift due to the 1000 Rads(Si) exposure. 34632 also

exhibits the least decrease in slope during the 1000 Rads(Si) exposure. The data for 33820 is confounded because due to practical time constraints, the medium dose radiation test was begun without the customary 72 hour minimum warm-up period. Thus, in the 0-40 Rads(Si) interval the slope per rad is 3.31×10^{-11} for 33820 versus 3.10×10^{-11} for 34632. In contrast, by the time the 500-1000 Rads(Si) interval was reached, df/f per rad for 33820 is considerably lower than that for 34632 (5.72×10^{-13} versus 2.04×10^{-12} respectively). The effect of inadequate warm-up for 33820 is also seen in the net shifts shown above for the complete 1000 Rads(Si) exposure. With an adequate warm-up period 33820 would have a net shift closer to that of 34744 (its companion from the same bar of quartz) rather than being closer to the net shift of 34632.

Table 3 compares the AT resonator drift rates, 10 second and 100 second Allan variances during and after the medium dose level irradiation. Resonators 34744 and 33820 have drift rates during irradiation which are factors of 2.20 and 2.31 higher than their post-irradiation drift values, respectively. Resonator 34632, the one with the greater aluminum content, exhibits a much greater differential when its during and after irradiation drift rates are compared as well as a difference in the sign of the drift rate. All three resonators exhibit considerable overlap in the ranges of their 10 and 100 second Allan variances when the during irradiation and after irradiation values are compared.

V. Recovery

Each of the three AT resonators tested in the medium dose environment exhibited net df/f shifts which

Table III
AT oscillator allan variances and drift rates during and after irradiation.

AT RESONATOR SERIAL NUMBER	OBSERVATION INTERVAL	DRIFT RATE* [$\Delta f/f$ PER DAY]	10 SECOND ALLAN VARIANCES [$\sigma(f)$]	100 SECOND ALLAN VARIANCES [$\sigma(f)$]
33820	DURING IRRADIATION	5.35×10^{-11}	$4 - 5 \times 10^{-12}$	$1 - 1.5 \times 10^{-12}$
	AFTER IRRADIATION	2.32×10^{-11}	$2.5 - 3.7 \times 10^{-12}$	$0.97 - 1.4 \times 10^{-12}$
34632	DURING IRRADIATION	1.88×10^{-10}	$0.8 - 2.5 \times 10^{-12}$	$0.86 - 2.8 \times 10^{-12}$
	AFTER IRRADIATION	-2.75×10^{-11}	$0.83 - 1.15 \times 10^{-12}$	$0.72 - 1.49 \times 10^{-12}$
34744	DURING IRRADIATION	7.74×10^{-11}	$0.91 - 1.08 \times 10^{-12}$	$0.78 - 1.38 \times 10^{-12}$
	AFTER IRRADIATION	3.52×10^{-11}	$0.66 - 1.37 \times 10^{-12}$	$0.65 - 1.32 \times 10^{-12}$

* DRIFT RATE DURING IRRADIATION WAS MEASURED OVER THE 500-1000 RADS (Si) INTERVAL.

were positive and on the order of a few parts in 10^9 . When the radiation stopped on each of these resonators, negative df/f shifts occurred temporarily. For the low aluminum content resonators, 33820 and 34744, these rebounds were small, $df/f = 6.8 \times 10^{-11}$ and 5.8×10^{-11} respectively, and of short duration, 5.5 and 2.1 hours respectively. For resonator 34632, the negative df/f recovery was $df/f = 8.5 \times 10^{-10}$ and took place over 41 hours. The lower aluminum content resonators have recoveries of only 2-4% of their positive df/f shifts due to radiation while 34632 with the higher aluminum content has a recovery of about 20% before a representative post-irradiation drift rate is resumed. Figures 6, 7 and 8 show the recoveries of the three AT resonators. The least square functional fits to the data are discussed in Section VI.

We have not subtracted either the pre-irradiation or post-irradiation drift rates from the df/f performance during radiation because these two drift rates are not always the same. Although 34632 had comparable pre- and post-irradiation drift rates (both negative), 33820 suffered from lack of adequate warm-up, as previously mentioned, and 34744 exhibited a substantial reduction in drift rate after irradiation.

VI. Least Square Geometrical Functional Fits of Oscillator Radiation Test Data

The frequency shifts (df/f) versus total accumulated dose (R [Rads(Si)]) as shown in Figures 3 through 5 for the three AT oscillators were analyzed with the aid of a program written for a Programmable TI Calculator and later modified for use on an AT&T 6300 Computer. The results of these fits were also verified against a commercially available curve fitting program, which confirmed our previous findings (reference 8). A Least

Square Geometrical (LSG) fit was performed for each of the curves shown in Figures 3-5. The general mathematical aspects of LSG fits are discussed in a variety of texts (References 9, 10 and 11). The LSG fits are all of the form:

$$[df/f(R)] = AR^B \quad (1)$$

where R is the accumulated radiation dose [Rads(Si)] and $df/f(R)$ is unitless. Table IV summarizes the values of A and B for each of the fits together with the coefficient of correlation of the fits.

The correlation coefficient in the last column of Table IV is a measure of the degree of closeness of the geometrical relationship between the two variables [$df/f(R)$] and R . As discussed in Section V, resonator 33820 suffered from inadequate warm-up, which is reflected in Table IV by the difference between the values of the coefficient A of the fits of resonators 33820 and 34744. However, the value of coefficients B are within 2% of each other. The value of the coefficient B for resonator 34632 with an AL content of 8.3 ppm is 20% greater when compared with the low AL content resonators. A modest interpretation might be that the coefficients for the functional fits may be similar for resonators with the same AL-impurity content for exposure to gamma radiation over a range from 0 to 1000 Rads(Si). This, in our opinion, warrants further investigation using a larger data base of resonators with different and equal AL-impurity contents.

As was previously mentioned in section V of this paper, figures 6 through 8 show the recoveries of the AT oscillators as a function of time. Using mathematical techniques analogous to the ones discussed in the previous paragraph, we calculated the LSG fits for each of the recovery curves shown in the aforementioned figures. The values of the LSG coefficients and the correlation factors

Table IV
Coefficients for least square geometrical fits for oscillator radiation tests function: $[df/f(R)] = AR^B$.

RESONATOR SERIAL NUMBER	AL-IMPURITY CONTENT (ppm)	$A[\times 10^{-9}]$	B	COEFFICIENT OF CORRELATION (r)
33820	0.79	0.471	0.335	0.976
34744	0.79	0.107	0.340	0.998
34632	8.3	0.288	0.408	0.992

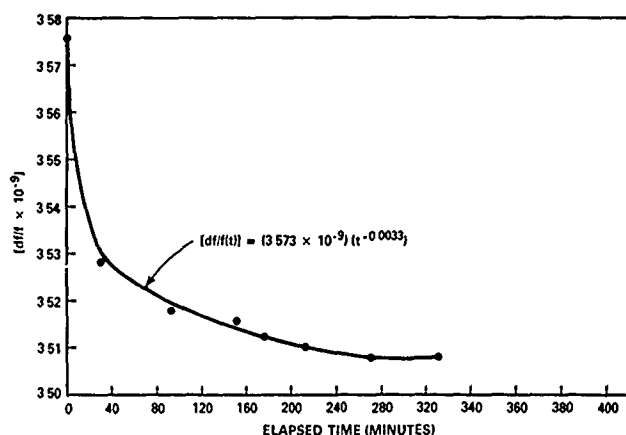


Fig. 6 At oscillator 33820 radiation recovery data.

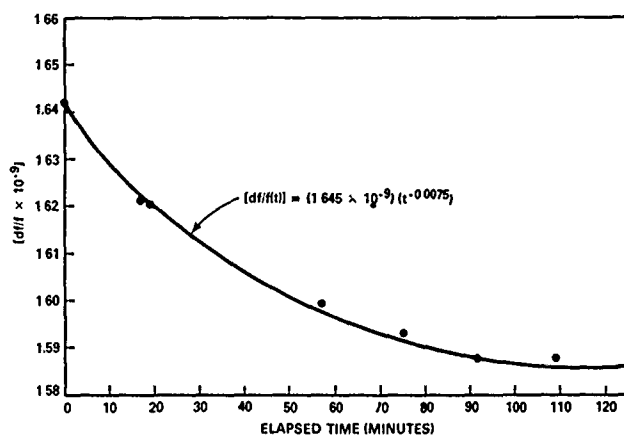


Fig. 8 At oscillator 34744 radiation recovery data.

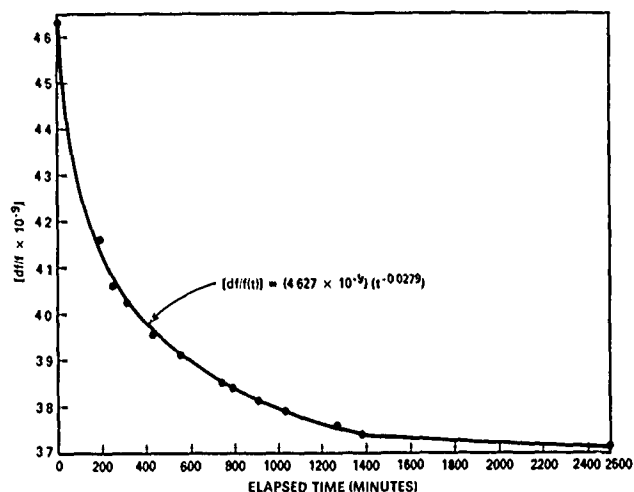


Fig. 7 At oscillator 34632 radiation recovery data.

are shown in Table V. From this table it may be seen that the LSG fit is of the type:

$$[df/f(t)] = Ct^D \quad (2)$$

where $[df/f(t)]$ is dimensionless and t is in minutes. Typically, as figures 6 through 8 suggest, the exponent D takes on negative values. Furthermore, the magnitude of this exponent apparently depends on the AL content of the resonator blank. We are of the opinion that expanding the data base of radiation tests on resonators might confirm this theory.

VII. Conclusions and Summary

1. As reported previously (Reference 1), we have confirmed that no correlation exists between the normalized frequency shift, df/f , and resonator

aluminum impurity content in the low radiation dose range of 0-4 Rads(Si), presumably due to the confounding of resonator surface effects.

2. Three AT cut resonators have now been tested to the medium dose level of 1000 Rads(Si). One of these resonators had an aluminum impurity content an order of magnitude greater than those of the other two resonators. All three resonators exhibited similar response profiles under the 1000 Rads(Si) exposure with df/f being positive and with df/f versus dose response saturating as the dose increased.

3. At the 1000 Rads(Si) total dose level the net df/f shifts are correlated with aluminum impurity content. In the 500-1000 Rads(Si) interval the slopes of df/f versus dose are correlated with aluminum impurity content. Both greater net shifts and greater slopes occur with the resonator (34632) of greater aluminum content. Quantitatively, the 500-1000 Rads(Si) interval slope of the resonator with a factor of ten greater AL-impurity content was a factor of 2-3 bigger than the slopes of the two resonators with the smaller AL content.

4. All recoveries after irradiation for the three AT cut quartz crystal resonators featured negative df/f shifts or "rebounds" from the shifts due to irradiation. The resonator (34632) with the order of magnitude greater aluminum content had a rebound amplitude and a recovery time which were an order of magnitude greater than those for the lower AL content resonators (20% of the radiation induced offset versus 2% and 41 hours versus 2-5 hours, respectively).

5. The generation of stable frequencies from quartz crystals is based on the piezo electric effect (Reference 12). The direct piezo electric effect may be represented by:

Table V
Coefficients for least square geometrical fits for oscillator recovery function: $[df/f(t)] = Ct^D$.

RESONATOR SERIAL NUMBER	AL-IMPURITY CONTENT (ppm)	$C(\times 10^{-9})$	D	COEFFICIENT OF CORRELATION (r)
33820	0.79	3.573	-0.0033	0.983
34744	0.79	1.645	-0.0075	0.975
34632	8.3	4.627	-0.0279	0.902

$$\sigma_{ij} = \sum_{k,l} c_{ijkl} u_{kl} - \sum_k e_{kij} E_k \quad (3)$$

where σ_{ij} is the second rank stress tensor, c_{ijkl} is the fourth rank elastic stiffness tensor, u_{kl} is the second rank strain tensor, e_{kij} is the third rank piezo electric tensor and E_k is the electric field vector (the indices i, j and k may take on values of 1, 2 and 3). The converse piezo electric effect is expressed by:

$$D_i = \sum_j \epsilon_{ij} E_j + \sum_{j,k} e_{ijk} u_{jk} \quad (4)$$

where D_i is the electric displacement vector and ϵ_{ij} is the electric permittivity (Reference 13).

As we observed in our experiment, the AL impurities affect the frequency stability of a quartz crystal resonator more when the impurity count is high. As may be observed from the constitutive piezo electric equations (3 and 4), not only can the AL ions effect the magnitude of the elastic stiffnesses (c_{ijkl}) but their net electric charge or interstitial charge compensator may also contribute to variations of the electric field vector (E_k) within the quartz crystal. As stated by King and Koehler (Reference 14), it is not readily apparent how the radiation effects specific elastic stiffnesses (c_{ijkl}) of the quartz crystal and/or the net electric field within the crystal. In the experiment described in this paper we observed that the greater the crystal's AL impurity content the greater its frequency shift and recovery period (t) for total accumulated doses of up to 1000 Rads(Si). We are of the opinion that further experimentations with quartz crystal resonators with different AL-impurity content, covering a larger range may confirm our initial observations.

VIII. References

1. J. R. Norton, J. M. Cloeren, and J. J. Suter, "Results from Gamma Ray and Proton Beam Radiation Testing of Quartz Resonators," IEEE, Transactions on Nuclear Science, NS-31, p. 1230 (1984).
2. J. J. Suter and R. H. Maurer, "Low and High Dose Photon Irradiation of Quartz Crystal Resonators," to be published in the Proceedings of the 17th Annual PTTI Conference, Washington, DC, 1985.
3. T. M. Flanagan, "Hardness Assurance in Quartz Crystal Resonators," IEEE Transactions on Nuclear Science, NS-21, p. 390 (1974).
4. P. Pellegrini, e.a., "Steady State and Transient Radiation Effects in Precision Quartz Oscillators," IEEE Transactions on Nuclear Science, NS-25, p. 1267 (1978).
5. H. G. Lipson, e.a., "Radiation Effects in Swept Premium - Q Quartz Material Resonators and Oscillators," Proceedings of the 33rd Annual Frequency Control Symposium, p. 122 (1979).
6. F. Euler, e.a., "Radiation Effects in Quartz Oscillators, Resonators and Materials," Proceedings of the 34th Annual Frequency Control Symposium, p. 72 (1980).
7. L. E. Halliburton, "Report on Evaluation of Synthetic Quartz," Department of Physics, Oklahoma State University Reports of 31 March 1981 and 28 April 1982.
8. Interactive Microware, Inc. Curve-Fitter-PC State College, PA (1984)
9. G. W. Snedecor and W. G. Cochran, "Statistical Methods," The Iowa State University Press, Ames Iowa (1973)
10. C. Daniel and S. Wood, "Fitting Equations to Data," J. Wiley, New York (1971)
11. D. C. Baird, "Experimentation: An Introduction to Measurement Theory and Experiment Design," Prentice-Hall, Englewood Cliffs, NJ (1962)
12. W. P. Mason, "Piezo Electric Crystals and Their Application to Ultrasonics," D. Van Nostrand Co., New York (1950)
13. J. F. Nye, "Physical Properties of Crystals," Clarendon Press, Oxford (1985)
14. E. A. Gerber; e.a., "Precision Frequency Control, Volume 1 Acoustic Resonators and Filters," Academic Press, Inc., New York (1985)

Ψ ANGLE IN A TRIPLY ROTATED CUT, DETERMINATION AND CONTROL

Henri Méricoux and Jean-François Darces

Laboratoire de Cristallographie et Synthèses Minérales
Faculté des Sciences et des Techniques
25030 Besançon Cedex - FranceSummary

Triply rotated cuts begin to be more and more demanded. The IEEE recommendations are suitable in preparing these cuts. After introducing Θ and Φ angles a third Ψ angle is needed.

In preparing a triply rotated cut there is an intermediate step we must absolutely control. Which is the Φ angle signs of the sides of the block already obtained. It is known that in a blank with parallel faces the orientation of the two faces differs only by the sign of the Φ angle. It is also known that the sign of the Φ and Ψ angles are linked. An error of the Φ angle sign induces an error on the Ψ angle sign.

In our laboratory using an X-ray diffraction goniometer we propose a method giving the right sign of the Φ angle and therefore a process to obtain a cut with a good Ψ angle. On the prepared block once on the goniometer we determine the Φ angle with the sign of the front side of the block. At the same time we know the ZZ' direction position on the block. This method requires knowledge of the scattering factors of the observed lattice planes and a test taking into account the plane observation order.

Key words : quartz, piezoelectricity, triply rotated cut, X ray diffraction.

Introduction

Triply rotated cuts begin to be more and more demanded. Particularly in the case of SAW devices where the most interesting orientations correspond to very general space directions.

Quartz is still the most used crystal among the piezoelectrical materials. Its symmetry allows some simplifications in the definition of the cut we use. These simplifications may be useful in the cut preparations.

IEEE Recommendations

Before introducing the triply rotated cut determination or control we must reintroduce the IEEE recommendations.¹

A Θ , Φ , Ψ cut is obtained from an hypothetical plate having the shape of the expected cut. The position of this plate in the XYZ axis system reference axis is very simple. In figure 1 we present a (Y,X) plate, the first letter, here Y, means that the thickness is along the Y axis, the second letter, here X, the length along the X axis.

A first rotation moves this plate. In figure 2 the rotation axis is the Z axis or the width of the plate. The angle is positive if the rotation is counterclockwise looking down the position end of the axis toward the origin. Here the sign of the rotation is negative and the plate may be (YXw) - Φ . The perpendicular to the plate is now Y'.

A second rotation moves the plate we have just had. In figure 3 the rotation axis is X' or the length of the plate ; Θ is positive. The plate is

labeled (YXw) - Φ /+ Θ . The perpendicular to the plate is now Y''. The third rotation around the thickness or the Y'' axis moves the plate to the expected position (YXwlt) - Φ /+ Θ /- Ψ as in figure 4.

In the case of quartz there is a very useful simplification we must introduce.

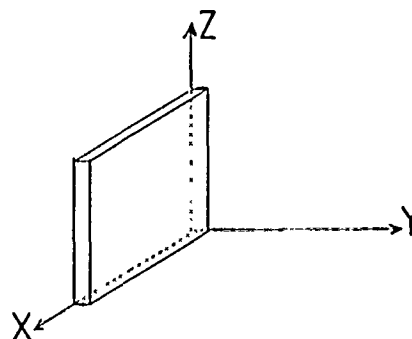
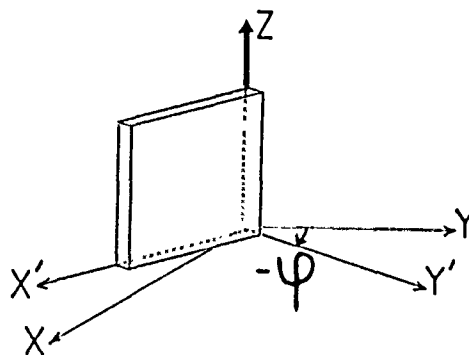
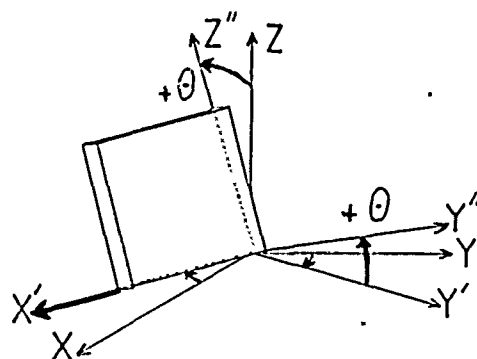
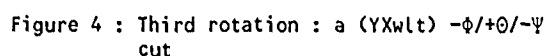
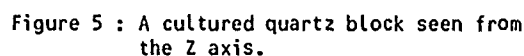


Figure 1 : Starting plate : a (YX) cut

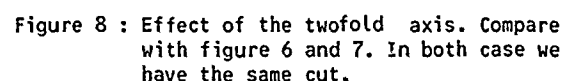
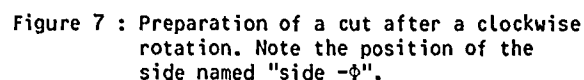
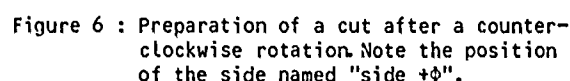
Figure 2 : First rotation : a (YXw) - Φ cutFigure 3 : Second rotation : a (YXwlt) - Φ /+ Θ cut



Instead of a general Θ, ϕ cut we introduce the simple case of a $\Theta = 0, \phi = +\phi$ cut. We may obtain it in a cultured quartz block represented in figure 5



The X axis is a twofold axis in quartz. The second cut $\phi = 0$, $\phi = -\phi$ may be repeated by this symmetry axis. In figure 8 we have the two plates linked by the twofold symmetry axis. They are identical we may recognize that the new plate is also identical with the $\phi = 0$, $\phi = +\phi$ cut.



Then the cut $\Theta = 0$, $\phi = +\phi$ and the cut $\Theta = 0$, $\phi = -\phi$ are identical. They both have a side named "side $+\phi$ " and a side named "side $-\phi$ ". We have the same cut using two different definitions. So we do not need to introduce the sign of the ϕ angle. A doubly rotated cut may be named Θ , $|\phi|$.

This is a very useful property: when we prepare a doubly rotated cut we do not have to take into account the sign of the ϕ angle. We just have to cut in such a way that the volume of the block is well employed. When the cuts are prepared we have, if necessary, to recognize which side is which.

Remark I : If we introduce a dissymmetry on the resonator with special electrodes or with a convexity on one of the two sides, then we must indicate which side is concerned. This indication will be a much better description of the resonator if we indicate the side concerned by the electrodes: for instance we may say a : Θ , $|\phi|$ S.A.W. resonator, with electrodes on "side $+\phi$ ".

Remark II : This property appears with enantiomorphous crystals belonging to 222, 32, 422, 622 symmetry groups and also 42m where a twofold axis is parallel to the X axis.

With crystal belonging to m, 2mm, 3m, 4mm, 6mm groups the sign of ϕ is of no importance.

With the other piezoelectrical groups there is no relation as those we have presented.

Determination of a triply rotated cut

A triply rotated cut is obtained on a doubly rotated cut by rotating around the Y'' axis. The rotation angle is Ψ . The sign of Ψ is defined using the positive end of the rotation axis.

We have shown that with quartz there is only one doubly rotated cut : Θ , $|\phi|$. Now we show there are only two triply rotated cuts, one with each possible sign of the Ψ angle.

For that the easiest way is to consider a very simple example. In figure 9 a Escher's drawing² represents a plane without symmetry as a doubly rotated cut.

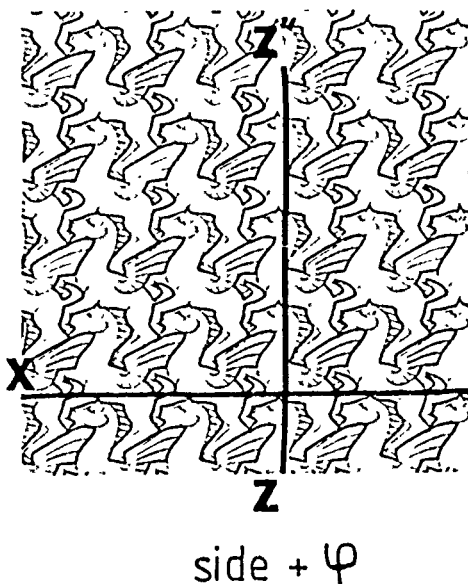


Figure 9 : Escher's drawing as an image of a doubly rotated cut. The side $+\phi$ is up.

We suppose we know where the Z'' axis is located on this plane and that the "side $+\phi$ " is up. We suppose also we want a rectangular plate with the width defined by the angle $-\Psi$. From the Z'' axis, a clockwise rotation, (figure 10), gives us the direction of the edge of the plate. We cut the rectangular cut, (figure 11) and we obtain the expected cut (figure 12).

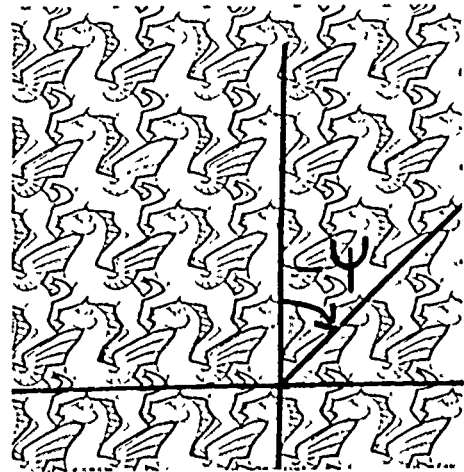


Figure 10 : Preparation of the triply rotated cut.

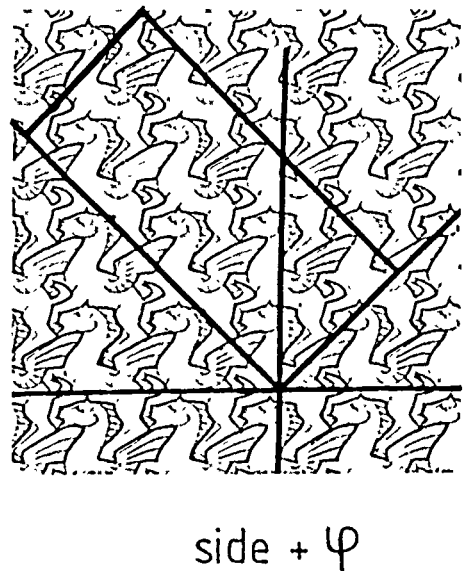
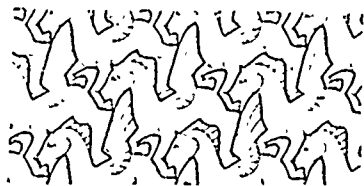


Figure 11 : Preparation of the rectangular plate.

The physical properties of this resonator may be symbolized by the line of flying Pegasus.

If we turn the plate upside down, the Pegasus look the other side of the sheet (figure 13).

The same clockwise rotation, (figure 14), and the same preparation (figure 15) give us a cut we recognize as an unexpected cut. The physical properties of the resonator will be surely different (figure 16).



the expected cut

Figure 12 : The line of flying Pegasus is supposed to be a physical property Compare with figure 16.

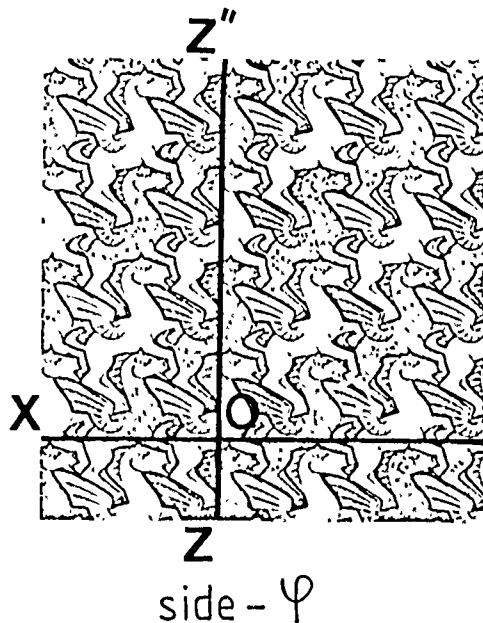
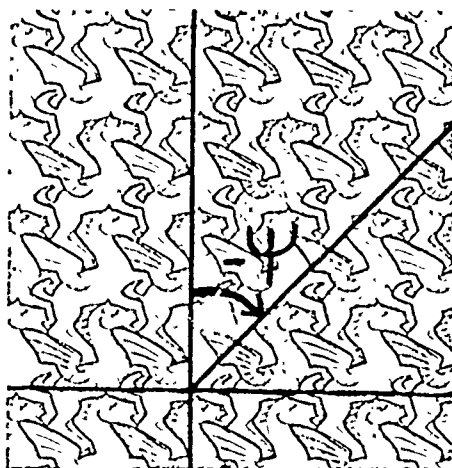
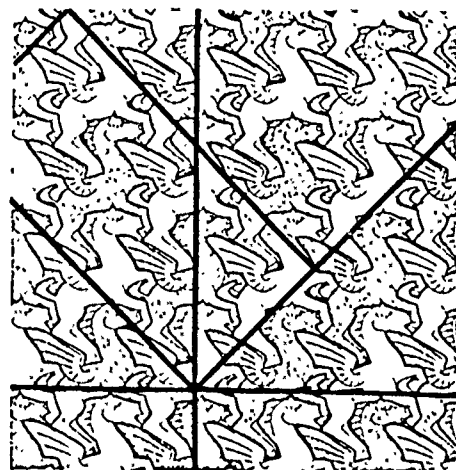


Figure 13 : The plate is upside down



side - ψ

Figure 14 : Preparation of the triply rotated cut.



side - ψ

Figure 15 : Preparation of the rectangular plate.



the unexpected cut

Figure 16 : The "side + ϕ " is down. Compare with figure 12.

We can immediately foresee we would have had the first position if we had associated a + ψ rotation (counterclockwise) with the side named "side - ϕ ".

We have the same result if we write θ , + ϕ , - ψ or θ , - ϕ , + ψ .

The simplest way is to present only two triply rotated cuts.

One : θ , + ϕ , + ψ

The other : θ , + ϕ , - ψ

The situation θ , - ϕ , - ψ is equivalent to θ , + ϕ , + ψ and θ , - ϕ , + ψ is equivalent to θ , + ϕ , - ψ .

These results help us. We have no difficulty with the definition of the positive end of the Y'' axis used in the rotation definition ; we may say : if the ϕ angle sign is +, the Y'' positive end is near the starting Y axis end, if the ϕ angle sign is -, the Y'' negative end is near the starting Y axis end.

Once this definition accepted we may use easily the IEEE recommendations. On a experimental point of view when we think of or we prepare a triply rotated cut we just need to know where the side named "side + ϕ " is located. This gives us the position of the positive end of the axis needed for the ψ angle sign. All sources of error are avoided

because we have now no ambiguity on the procedure needed for the preparation.

We think we must definitely give up the use of the +X axis position. The position of the +X axis is of no help, and it may induce errors. Because this axis is a twofold axis, it does not move when the side + or - ϕ go upside down.

PREPARATION AND CONTROL

Now the preparation and the control of a triply rotated cut is very easy. When it is possible, the best thing is to place a mark on the plate giving the name "side + ϕ or - ϕ " on one of the two sides of a plate. If there is a doubt or if the marks are gone we may use a piezoelectrical test or an X ray determination. With the X ray determination it is very useful if the goniometer permits the observation of any lattice planes. We need at least three lattice planes ; but according to the orientation these planes may be different.

When we observe the (hkl) lattice planes the vector $\vec{s}_{(hkl)}$ perpendicular to the plane may be defined by two angles ρ , μ . ρ : angle between $\vec{s}_{(hkl)}$ and \vec{Ng} perpendicular to the plate ; μ : angle between the $\vec{s}_{(hkl)}$ projection on the plate and the edge we need for the definition of the triply rotated cut.³ (figure 17). These two angles are very easily measured on the goniometer we use in our laboratory.⁴

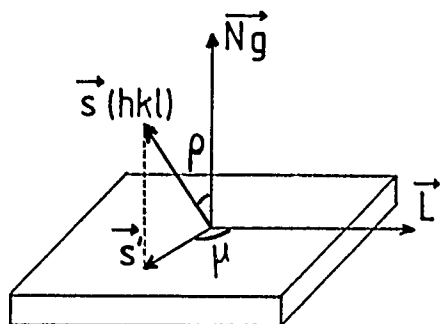


Figure 17 : A lattice plane (hkl) is located on the plate by two angles : ρ , μ .

Using these ρ measurements we know the position of \vec{Ng} in the axis system constructed with three observed $\vec{s}_{(hkl)}$ directions. A mathematical conversion gives the position with the θ and ϕ angles in the IEEE reference axis system, (figure 18). This result gives a bulk orientation, we have to determine the sign of the θ and ϕ angle. The sign of θ is known using the scattering factor associated with some of the observed lattice planes. The sign of ϕ is known using a comparison between a calculated and observed appearance order of the reflexion on the (hkl) lattice planes.⁵ Both of these results give the name of the side you are observing on the goniometer. Then the position of the Z'' axis may be calculated from the μ angle associated to the hkl plane.

Generally the Z'' axis may be directly seen using the (00.6) lattice plane. The $\vec{s}_{(00.6)}$ direction is exactly the Z axis in the IEEE reference axes. This (00.6) is seen if $|\phi|$ is greater than $31^\circ 3'$ using a Copper X ray tube. With a Cobalt tube this limit is lowered to 7° .

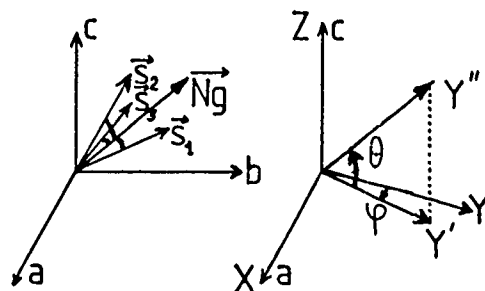


Figure 18 : The orientation expressed in a experimental axis system (left) is now given in the IEEE reference axis system (right).

If we prepare a cut we measure the angle between the Z'' axis and an edge of the observed doubly rotated cut. Then we calculate the difference giving the good Ψ angle with its sign. If it is a control we directly observe the absolute value of the Ψ angle. In both situations we take into account the position of the "side + ϕ " for the determination of the sign of the Ψ angle.

Conclusion

With quartz we may use a very useful simplification because there is only one doubly rotated cut named θ , $|\phi|$ and two triply rotated cuts θ , + ϕ , + ψ and θ , + ϕ , - ψ .

The knowledge of the position of the side named "side + ϕ " simplifies the preparation of a triply rotated cut and avoids confusion. When we prepare or control a cut by X ray diffraction we must know where the side named "side + ϕ " is on the plate and where the projection of the (00.6) lattice plane is located. This projection may be observed or calculated.

The IEEE recommendations are still useful. But in the case of quartz or crystal analog to it, some new explanations taking into account these results must be done to avoid confusion.

Acknowledgements

We would like to express our thanks to the D.R.E.T. (Direction des Recherches Etudes et Techniques) of the Ministry of Defense who sponsored this research.

References

- 1) IEEE Standard on Piezoelectricity, The Institute of Electrical and Electronics Engineers, Inc. 1978.
- 2) C.H. Macgillavry, Symmetry aspects of the M.C. Escher's periodic drawing, 1976, Bohn, Scheltema, Holkema, Utrecht.
- 3) Problèmes posés par la détermination de l'orientation absolue d'une lame monocristalline : application au quartz. Thèse d'Etat, 1984, J.F. Darces.
- 4) J.F. Darces, H. Mérigoux, Final X ray control of the orientation of round or rectangular slides for industrial purposes. Proceedings of the 32nd Annual Symposium on Frequency Control, 1978, p. 304-309.
- 5) J.F. Darces, J. Lamboley and H. Mérigoux, Dissymmetry used for ϕ , θ angle sign of a piezoelectric crystal blank by a non destructive method. Ferroelectrics 1982, vol. 40, p. 245-248.

DYNAMIC PERMITTIVITIES AND RESISTIVITIES OF THE EQUIVALENT NETWORK REPRESENTING PLATE RESONATORS

Arthur Ballato, Elizabeth Hatch,
Muhammad Mizan, and Theodore J. LukaszekUS Army Electronics Technology & Devices Laboratory
Fort Monmouth, New Jersey 07703-5000

ABSTRACT

A recent sequence of papers investigated the piezo coupling factors, viscosities, and time constants of singly and doubly rotated cuts of quartz plates executing thickness modes. These physics-type coefficients were briefly described in terms of the parameters of the equivalent electrical network characterizing a resonator. In this paper we use the representation to arrive at engineering-type coefficients that are pertinent to practical designs utilizing the Butterworth-Van Dyke equivalent circuit. Specifically calculated are the motional capacitance constant Γ_1 and the motional resistance constant P_1 as functions of PHI and THETA orientation angles of doubly rotated quartz plates driven by thickness and lateral fields. All three thickness modes are considered, and plate orientations having practically useful behavior are treated.

INTRODUCTION

High frequency bulk wave piezoelectric resonators use thickness modes of vibration almost exclusively. These, in turn, are nearly always driven by impressed voltages across electrodes situated on the plate major surfaces. This type of drive is known as thickness excitation (TE). Electrically, the resonators are normally represented by the traditional Butterworth-Van Dyke (BVD) circuit consisting of a shunt capacitor C_0 in parallel with a series R_1, C_1, L_1 string. Strictly speaking, the series arm for TE also includes a negative C_0 . This is usually omitted for simplicity, but can be important for high coupling materials; it changes the capacitance ratio by one unit.

The static capacitance C_0 arises from the dielectric permittivity of the crystal and the electrode arrangement; for TE the parallel-plate capacitance formula gives

$$C_0 = \epsilon * A/t. \quad (1)$$

The motional capacitance C_1 and the static capacitance C_0 are related by the piezo coupling factor k :

$$C_0/C_1 = r = \pi^2/(8k^2). \quad (2)$$

Because k is a constant for a given crystal, cut, and mode of motion, C_1 has the same geometrical dependence as C_0 . One can thus write

$$C_1 = \Gamma_1 * A/t, \quad (3)$$

where Γ_1 is the motional capacitance constant introduced by Bechmann [1-5]. It represents the permittivity of the fictitious element C_1 , and is equal

to

$$\Gamma_1 = 8 \epsilon k^2 / \pi^2. \quad (4)$$

Γ_1 may have any value > 0 , whereas ϵ cannot be less than the permittivity of free space.

From considerations of the BVD circuit, the $R_1 * C_1$ product is a constant called the motional time constant and denoted τ_1 ; it is calculated from the viscosity tensor [6]. Like k , it depends upon crystal, cut, and mode of motion. Because $R_1 * C_1$ is a constant, Eq. (3) can be used to find the equivalent resistivity of R_1 . This is denoted P_1 and equals:

$$P_1 = \pi^2 \tau_1 / (8 \epsilon k^2). \quad (5)$$

This motional resistance constant was also introduced by Bechmann [4].

LATERAL-FIELD EXCITATION

When the exciting electric field is along the plate major surfaces, it is referred to as lateral excitation (LE). Some of the earliest calculations of LE are also due to Bechmann [e.g., 7]; see [8] for an extended list of LE references. LE leads to a BVD circuit without the negative C_0 , i.e., the conventional BVD circuit is proper to LE. For lateral-field excitation, the electrode system usually employed bears more similarity to surface acoustic wave (SAW) interdigital transducers (IDT's) than parallel-plate capacitor electrodes. Nevertheless, the static capacitance C_0 is of the form in Eq. (1). In [8-12] lateral-field calculations are given; the capacitance ratio is again free of geometrical factors when it is assumed that the piezo current is purely lateral.

We are therefore able to introduce an LE Γ_1 and P_1 , representing the fictitious permittivity and resistivity of C_1 and R_1 , in exactly the same fashion as Γ_1 and P_1 for TE. The results are given in Figures 1 through 30 for singly and doubly rotated cuts of quartz. The orientational notation is that of the IEEE, and is discussed in Ref. [8]. From these curves, the motional capacitance and resistance of a plate vibrator may be determined when the electrode geometry is known. Both TE and LE curves are supplied. In the case of R_1 , only that component due to viscous friction is determinate from the P_1 listed.

ENERGY TRAPPING

The foregoing pertains to the simple thickness modes of plates, i.e., plates with a uniform lateral

distribution of motional activity. In practical cases, the nonuniform mode shape resulting from application of "Energy Trapping" rules modifies the values of Γ_1 and P_1 . The following TE examples are typical for AT-cut filter crystals:

Theoretical (uniform motion); $\tau_{A1} = 11.83$ femtosecond

GAMMA_1 (fundamental) = 249.7 femtofarad/meter
 RHO_1 = 0.0474 ohm-meter

GAMMA_1 (third) = 27.74 fF/m
 RHO_1 = 0.427 ohm-m

GAMMA_1 (fifth) = 9.988 fF/m
 RHO_1 = 1.185 ohm-m

GAMMA_1 decreases with harmonic number squared; since τ_{A1} is independent of harmonic, and is equal to the $\text{GAMMA}_1 * \text{RHO}_1$ product, RHO_1 increases with harmonic squared.

Experimental (nonuniform motion); τ_{A1} function of harmonic

GAMMA_1 (fundamental) = 220. fF/m
 RHO_1 = 0.075 ohm-m; $\tau_{A1} = 165$ fs

GAMMA_1 (third) = 25. fF/m
 RHO_1 = 1.0 ohm-m; $\tau_{A1} = 25$ fs

GAMMA_1 (fifth) = 10. fF/m
 RHO_1 = 2.5 ohm-m; $\tau_{A1} = 25$ fs

Because τ_1 varies with harmonic, we infer that energy is being lost to the mounting structure; the modes are insufficiently quiet at the mounting points. Air damping in unevacuated units produces additional losses.

ELECTROELASTIC EFFECT REMARK

This effect is of current interest, both theoretically [13] and technologically [14]. It appears that obtaining a consistent set of tensor components is difficult from measurements of TE plates only; one must augment these with results from bars or rods, or use pulse-echo techniques. A consistent set may be determined solely from resonator plates providing that the TE measurements are supplemented with LE measurements.

REFERENCES

1. R. Bechmann
Contour modes of square plates excited piezoelectrically and determination of elastic and piezoelectric coefficients.
Proc. Phys. Soc. London, B, Vol. 64, 1951, 323-337.
2. R. Bechmann and P. L. Parsons
Piezoelectric behavior of partially plated square plates vibrating in contour modes.
Proc. Phys. Soc. London, B, Vol. 64, 1951, 706-712.
3. R. Bechmann
Contour modes of plates excited piezoelectrically and determination of elastic and piezoelectric coefficients.
4. R. Bechmann and S. Ayers
Mechanical and electrical behaviour of piezoelectric crystals (equivalent electrical circuit). Report No. 3 in Piezoelectricity, H. M. S. O., London, 1957.
5. R. Bechmann
Piezoelektrisch erregte Eigenschwingungen von Platen und Staeben und dynamische Bestimmung der elastischen und piezoelektrischen Konstanten. Archiv der Elektr. Uebertragung, Vol. 8, 1954, 481-490.
6. A. Ballato
Doubly rotated thickness mode plate vibrators. In Physical Acoustics, Vol. 13, (W. P. Mason and R. N. Thurston, eds.), 115-181. Academic, New York, 1977.
7. R. Bechmann
Ueber Dickenschwingungen piezoelektrischer Kristallplatten. Archiv der Elektr. Uebertragung, (Nachtrag), Vol. 7, 1953, 354-356.
8. A. Ballato, E. R. Hatch, M. Mizan, and T. J. Lukaszek
Lateral field equivalent networks and piezocoupling factors of quartz plates driven in simple thickness modes. IEEE Trans. Ultrason., Ferro., & Freq. Control, Vol. UFFC-33, July 1986, 385-392.
9. E. R. Hatch and A. Ballato
Lateral-field excitation of quartz plates. IEEE Ultrasonics Symp. Proc., 1983, 512-514.
10. A. Ballato, E. R. Hatch, M. Mizan, B. H. T. Chai, R. S. Tilton, and T. J. Lukaszek
Lateral-field excitation of berlinite. Proc. 38th AFCS, May 1984, 206-224.
11. A. Ballato, E. R. Hatch, M. Mizan, T. Lukaszek, and R. Tilton
Simple thickness plate modes driven by lateral fields. Proc. 39th AFCS, May 1985, 462-472.
12. A. Ballato, E. R. Hatch, and T. Lukaszek
Acoustic viscosities and time constants of the simple thickness modes of quartz plates. IEEE Ultrasonics Symp. Proc., 1985, 1143-1146.
13. C. K. Hruska
On the disagreement in the order of magnitude of the electroelastic constants of alpha-quartz. These proceedings.
14. R. Brendel, J.-J. Gagnepain, and J. P. Aubry
Impurities migration study in quartz crystal resonators by using electroelastic effect. These proceedings.

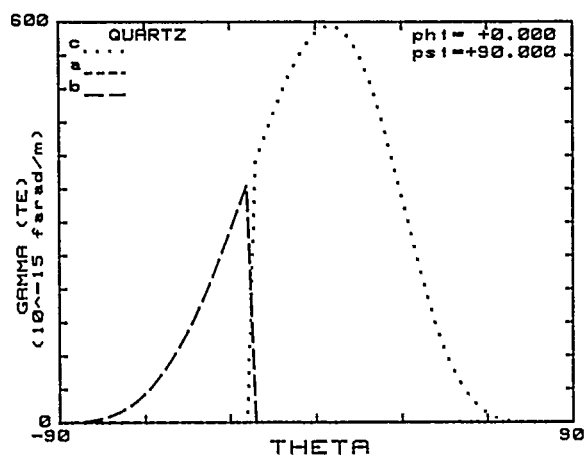


Figure 1. Gamma versus theta.
Phi = 0°, TE

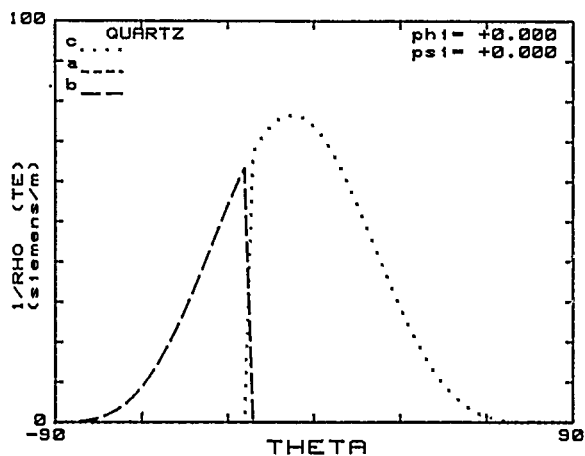


Figure 2. Reciprocal rho versus theta.
Phi = 0°, TE

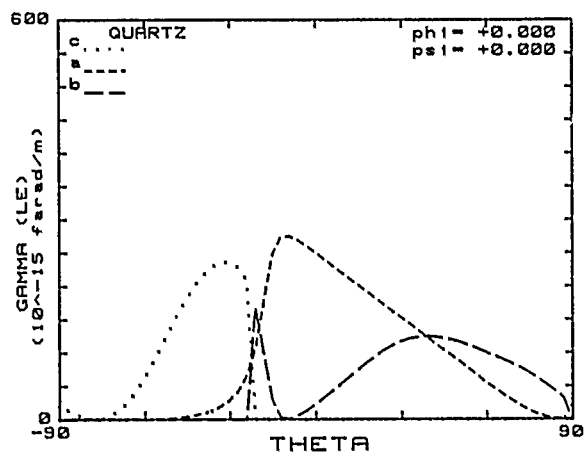


Figure 3. Gamma versus theta.
Phi = 0°, LE, psi = 0°.

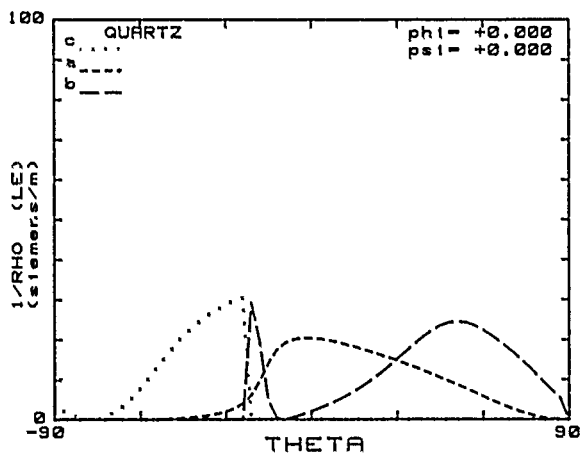


Figure 4. Reciprocal rho versus theta.
Phi = 0°, LE, psi = 0°.

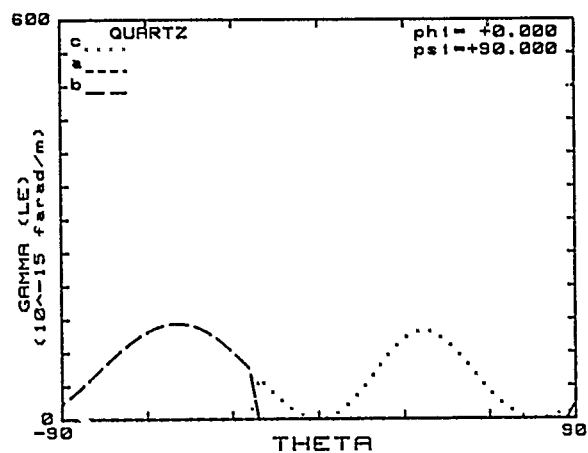


Figure 5. Gamma versus theta.
Phi = 0°, LE, psi = 90°

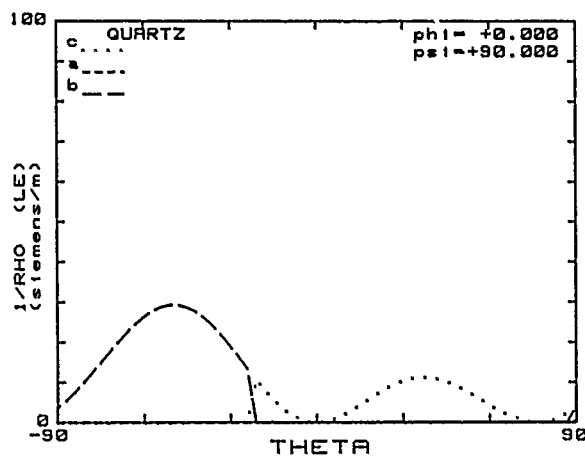


Figure 6. Reciprocal rho versus theta.
Phi = 0°, LE, psi = 90°.

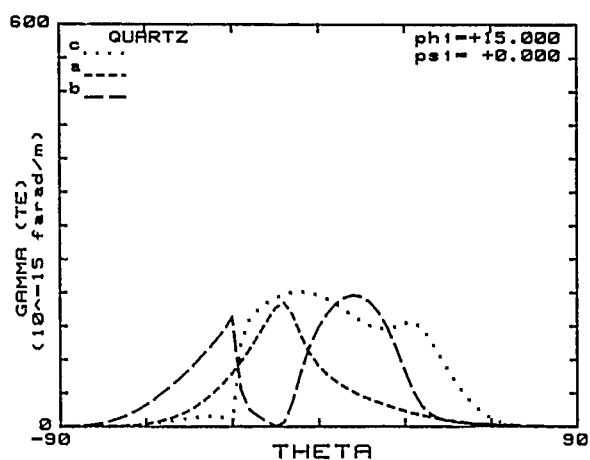


Figure 7. Gamma versus theta.
Phi = 15°, TE.

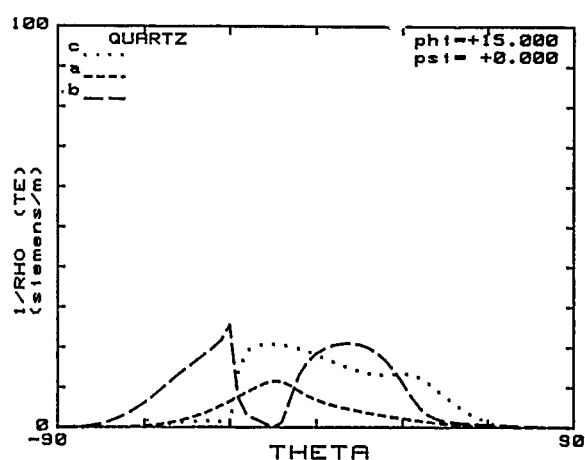


Figure 8. Reciprocal rho versus theta.
Phi = 15°, TE.

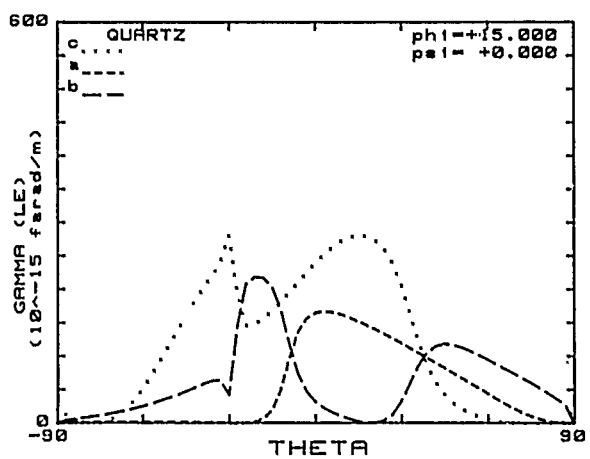


Figure 9. Gamma versus theta.
Phi = 15°, LE, psi = 0°.

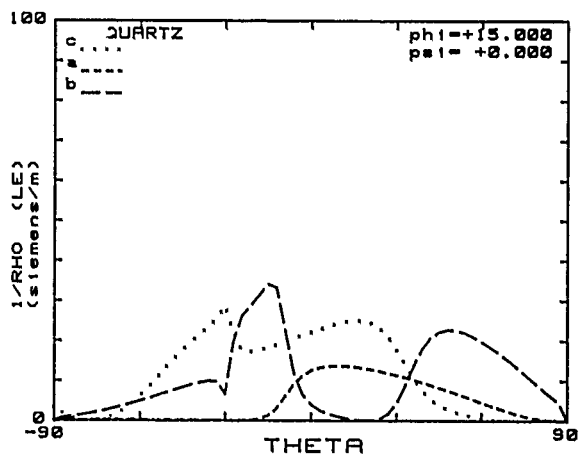


Figure 10. Reciprocal rho versus theta.
Phi = 15°, LE, psi = 0°.

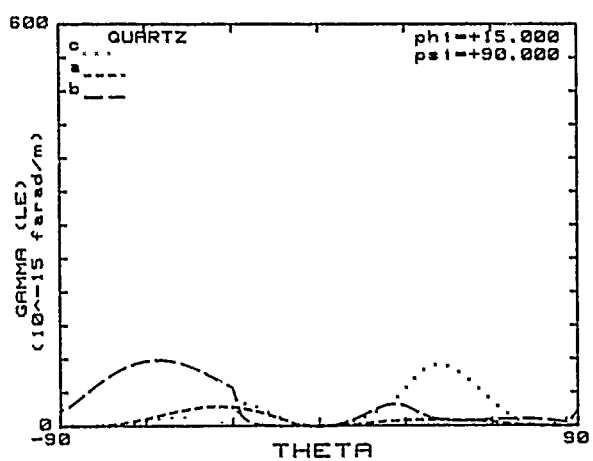


Figure 11. Gamma versus theta.
Phi = 15°, LE, psi = 90°.

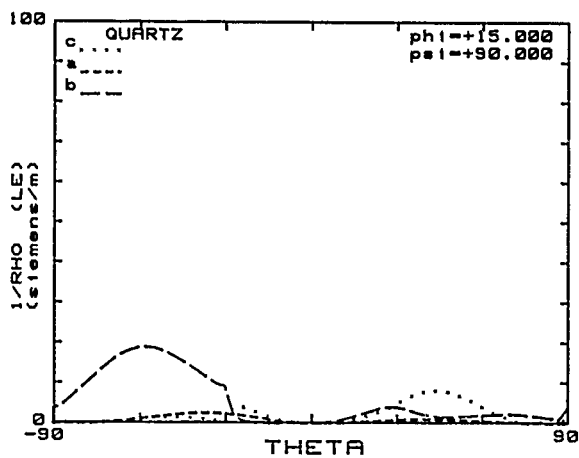


Figure 12. Reciprocal rho versus theta.
Phi = 15°, LE, psi = 90°.

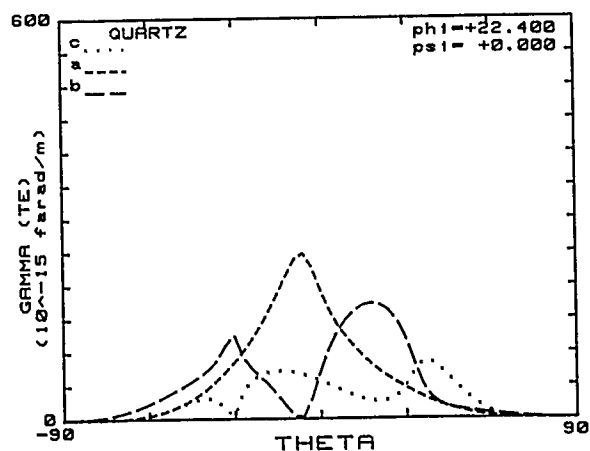


Figure 13. Gamma versus theta.
Phi = 22.4°, TE.

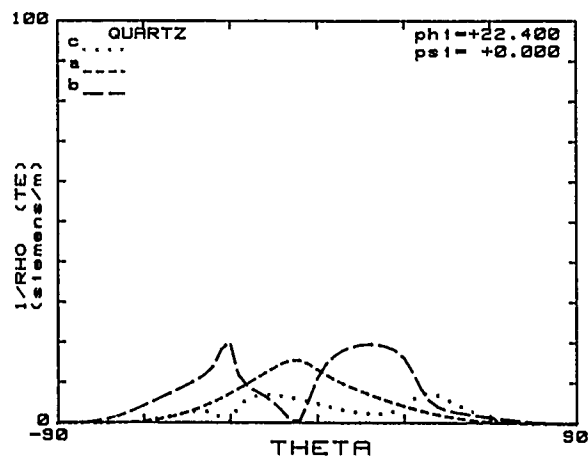


Figure 14. Reciprocal rho versus theta.
Phi = 22.4°, TE.

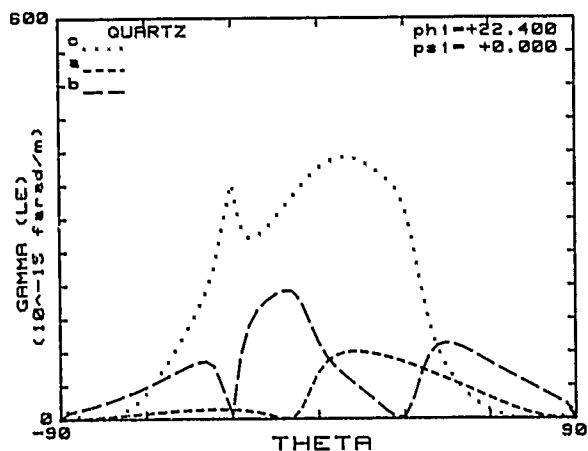


Figure 15. Gamma versus theta.
Phi = 22.4°, LE, psi = 0°.

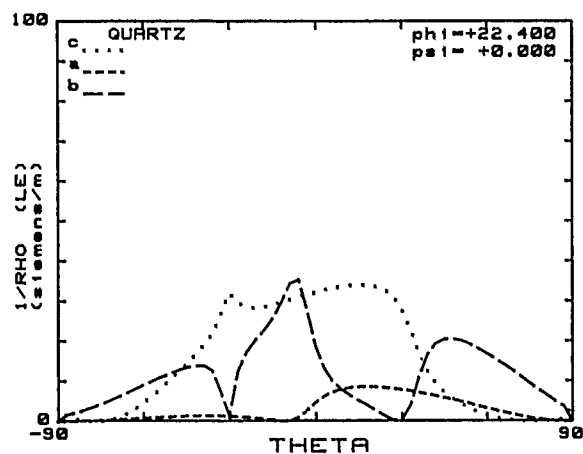


Figure 16. Reciprocal rho versus theta.
Phi = 22.4°, LE, psi = 0°.

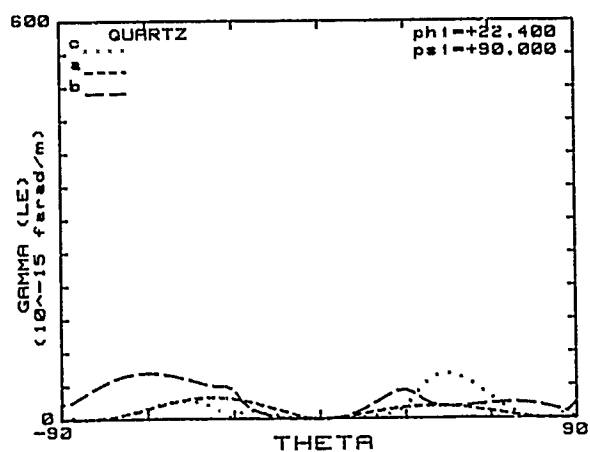


Figure 17. Gamma versus theta.
Phi = 22.4°, LE, psi = 90°.

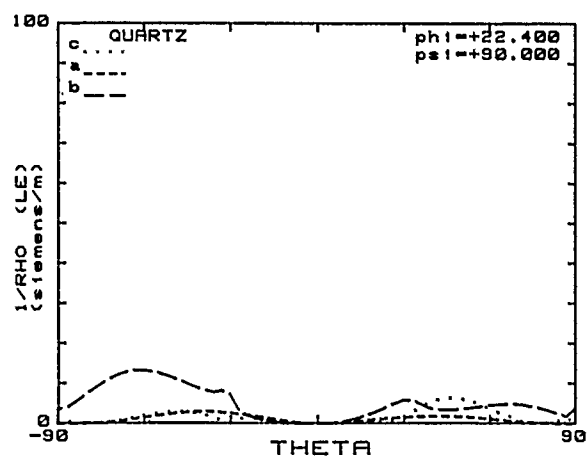


Figure 18. Reciprocal rho versus theta.
Phi = 22.4°, LE, psi = 90°.

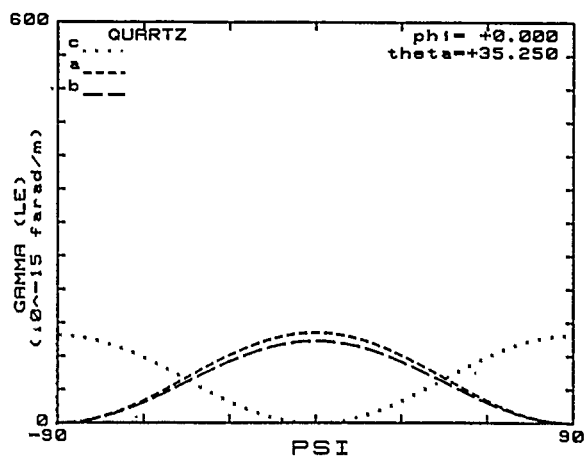


Figure 19. Gamma versus psi. AT cut.

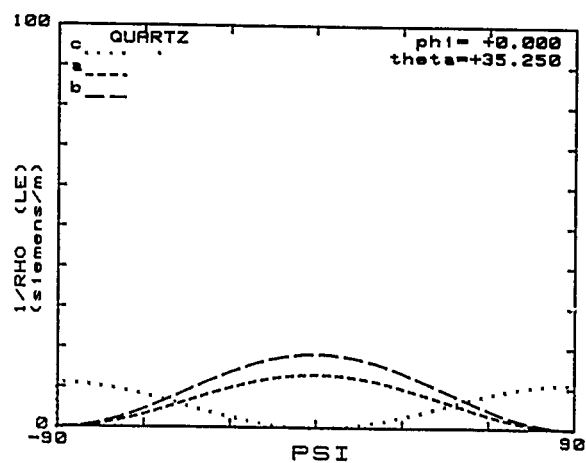


Figure 20. Reciprocal rho versus psi. AT cut.

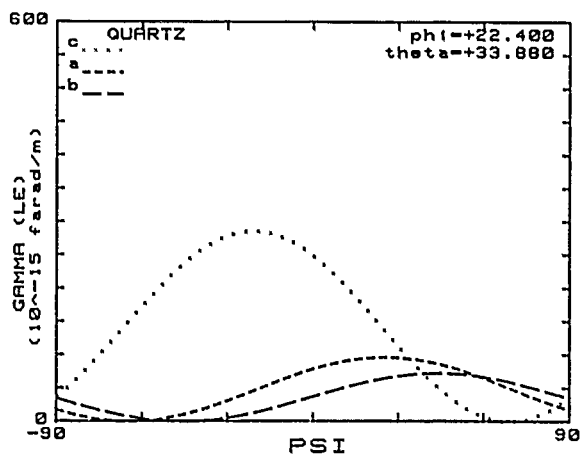


Figure 21. Gamma versus psi. SC cut.

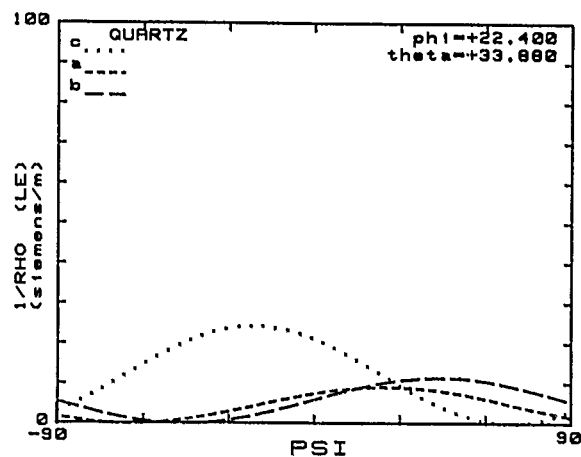


Figure 22. Reciprocal rho versus psi. SC cut.

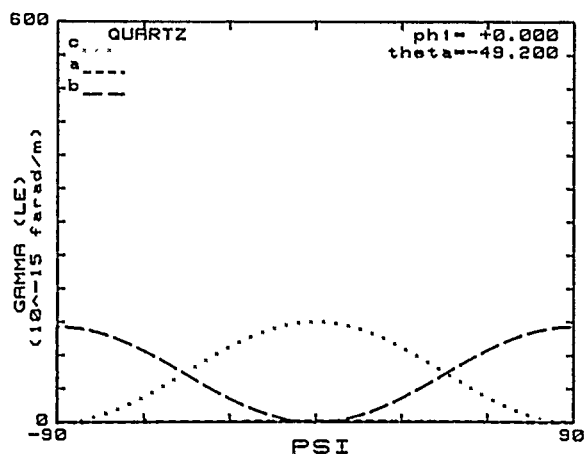


Figure 23. Gamma versus psi. BT cut.

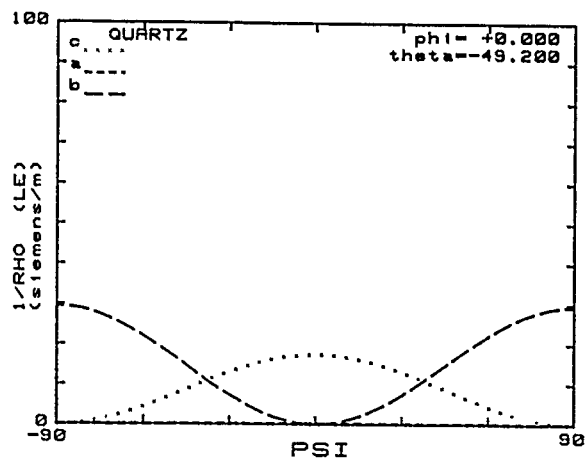


Figure 24. Reciprocal rho versus psi. BT cut.

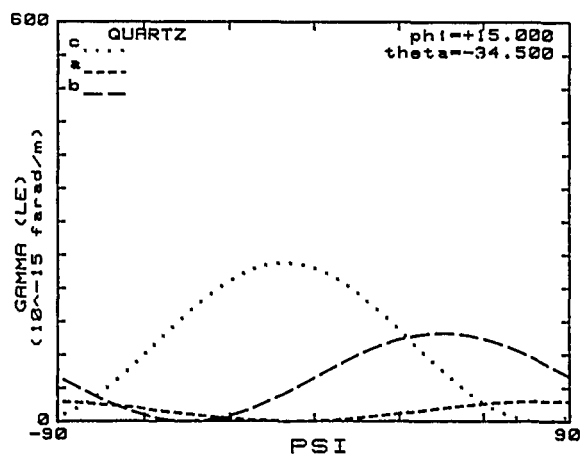


Figure 25. Gamma versus psi. RT cut.

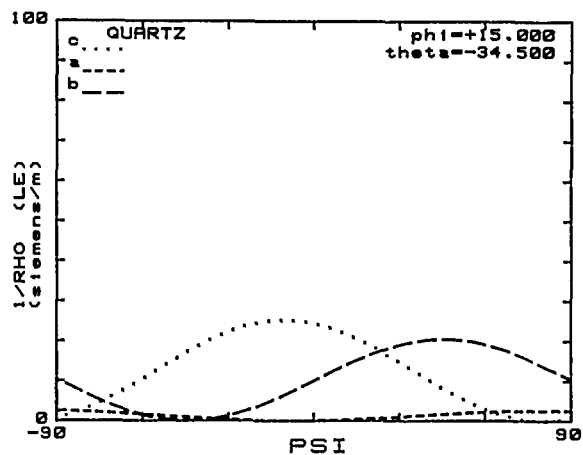


Figure 26. Reciprocal rho versus psi. RT cut.

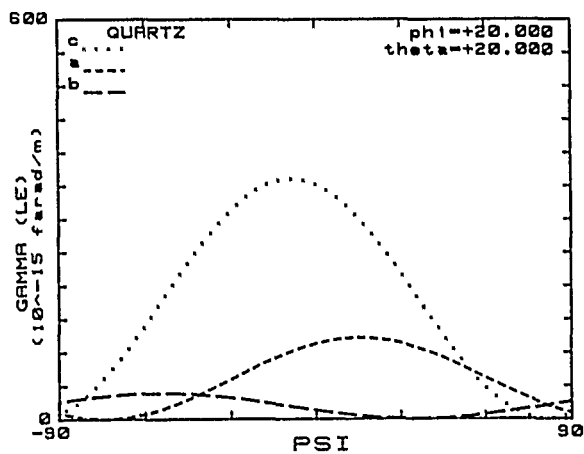


Figure 27. Gamma versus psi. NLSC cut.

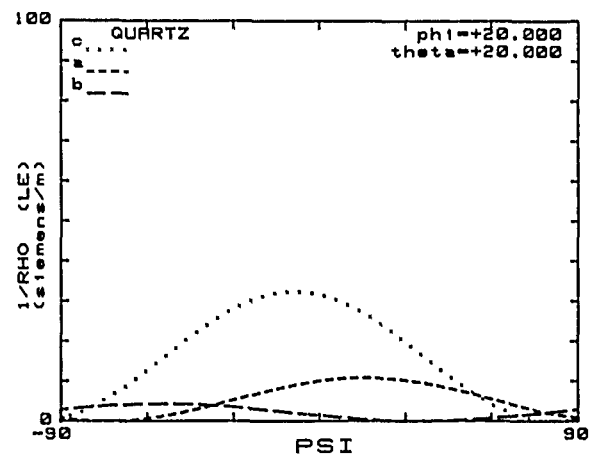


Figure 28. Reciprocal rho versus psi. NLSC cut.

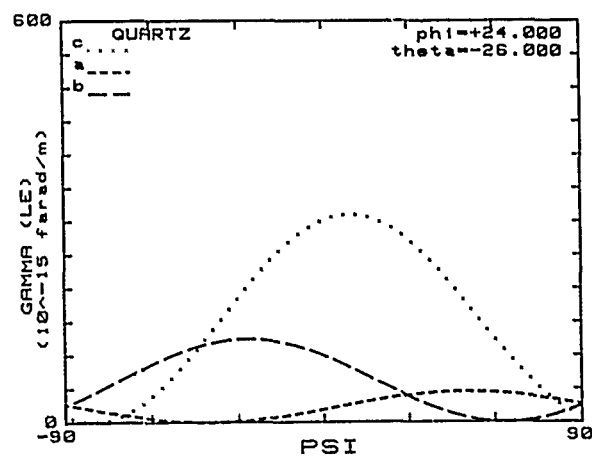


Figure 29. Gamma versus psi. AK cut.

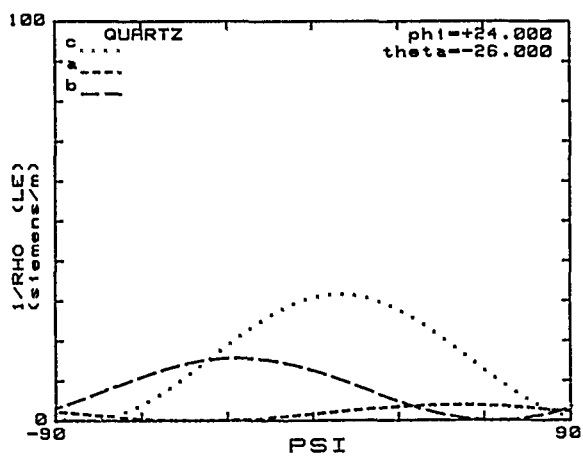


Figure 30. Reciprocal rho versus psi. AK cut.

INITIAL STRESS FIELD AND RESONANCE FREQUENCIES OF INCREMENTAL
VIBRATIONS IN CRYSTAL RESONATORS BY FINITE ELEMENT METHOD

P. C. Y. Lee and M. S. H. Tang

Department of Civil Engineering, Princeton University
Princeton, New Jersey 08544

Abstract

A finite element method for numerical solutions is employed to solve the six coupled two-dimensional equations of vibrations for doubly rotated crystal plates by recasting them into a variational boundary value problem. Based on these equations two computer programs have been developed for the numerical solutions.

The first program is for the purpose of calculating the initial stresses, strains, and displacements caused by the static applied forces and by the steady acceleration for the resonators with the most general crystal symmetry.

The second program is to predict the resonance frequencies and mode shapes of the free, incremental vibrations in the same crystal resonators.

Numerical results are obtained and compared with those from analytical or close form solutions for both isotropic and anisotropic plates and with experimental values.

Two-Dimensional Plate Equations

In the previous paper¹, two-dimensional equations of incremental vibrations superposed on finite, initial deformations were derived. When in stress-strain and strain-displacement relations of initial fields are linearized, the governing equations of initial fields and the governing equations of incremental fields without initial stresses reduce to the same form as Mindlin's first order plate equations.² These equations will be employed to predict initial stresses and strains due to applied forces or steady accelerations and to calculate the resonances and modes of free incremental vibrations in crystal resonators, and they are given below.

Stress equations of motion

$$\begin{aligned} T_{aj,a}^{(0)} + F_j^{(0)} + 2b\rho b_j^{(0)} &= 2b\rho u_j^{(0)} \\ T_{aj,a}^{(1)} - T_{2j}^{(0)} + F_j^{(1)} + \frac{2b^3}{3} \rho b_j^{(1)} &= \frac{2b^3}{3} \rho \ddot{u}_j^{(1)} \end{aligned} \quad (1)$$

where $a = 1, 3$ and $j = 1, 2, 3$.

Stress-strain relations

$$\begin{aligned} T_p^{(0)} &= 2b\kappa_{(p)}\kappa_{(q)} C_{pq} S_q^{(0)}, \quad p, q = 1, 2, \dots, 6 \\ T_p^{(1)} &= \frac{2b^3}{3} \bar{C}_{pq} S_q^{(1)}, \quad p, q = 1, 3, 4, 5, 6 \end{aligned} \quad (2)$$

where $\bar{C}_{pq} = C_{pq} - C_{2p}C_{2q}/C_{22}$

$$\begin{aligned} \kappa_{(p)} &= \pi/\sqrt{12}, \quad \text{when } p = 2, 4, 6 \\ &= 0 \quad \text{when } p = 1, 3, 5 \end{aligned}$$

Strain-displacement relations

$$\begin{aligned} S_1^{(0)} &= u_{1,1}^{(0)} \\ S_2^{(0)} &= u_2^{(1)} \\ S_3^{(0)} &= u_{3,3}^{(0)} \\ S_4^{(0)} &= u_{2,3}^{(0)} + u_3^{(1)} \\ S_5^{(0)} &= u_{3,1}^{(0)} + u_{1,3}^{(0)} \\ S_6^{(0)} &= u_{2,1}^{(0)} + u_1^{(1)} \end{aligned}$$

and

$$\begin{aligned} S_1^{(1)} &= u_{1,1}^{(1)} \\ S_3^{(1)} &= u_{3,3}^{(1)} \\ S_4^{(1)} &= u_{2,3}^{(1)} \\ S_5^{(1)} &= u_{3,1}^{(1)} + u_{1,3}^{(1)} \\ S_6^{(1)} &= u_{2,1}^{(1)} \end{aligned} \quad (3)$$

We see that (1) - (3) consists of a system of twenty-eight two-dimensional equations with twenty-eight unknowns. In the first order stress-strain relations (2)₂, $S_2^{(1)}$ is eliminated by substitution of

$$S_2^{(1)} = -\frac{1}{C_{22}} C_{2q} E_q^{(1)}, \quad q = 1, 3, 4, 5, 6 \quad (4)$$

which is, in turn, obtained by letter $T_2^{(1)} = 0$.

Variational Formulation

Equations (1) - (3) and their associated boundary conditions will be recasted into a variational equation which is a weaker and more suitable form for the application of finite element methods for numerical solutions.

We let $\bar{u}_1^{(n)}$ be the virtual displacements and

$$\begin{aligned} \bar{S}_{ij}^{(n)} &= \frac{1}{2} [\bar{u}_{j,i}^{(n)} + \bar{u}_{i,j}^{(n)} + n(\delta_{2i}\bar{u}_j^{(n+1)} \\ &\quad + \delta_{2j}\bar{u}_i^{(n+1)})] \end{aligned} \quad (5)$$

the virtual strains, where $n = 0, 1$.

Multiplying (1)₁ by $\bar{u}_j^{(0)}$ and (1)₂ by $\bar{u}_j^{(1)}$ and integrating the sum of the result over the plate area A, we have

$$\int_A (T_{aj,a}^{(0)} + F_j^{(0)} + 2b\rho b_j^{(0)} - 2b\rho u_j^{(0)}) \bar{u}_j^{(0)} dA + \int_A (T_{aj,a}^{(1)} - T_{2j}^{(0)} + F_j^{(1)} + \frac{2b^3}{3} \rho b_j^{(1)}) - \frac{2b^3}{3} \rho u_j^{(1)} \bar{u}_j^{(1)} dA = 0 \quad (6)$$

By applying the two-dimensional divergence theorem, and the use of (2) and (5) (6) is transformed to

$$\begin{aligned} & \int_A (2b \bar{S}_p^{(0)} \kappa_{(p)} \kappa_{(q)} c_{pq} \bar{S}_q^{(0)} + \frac{2b^3}{3} \bar{S}_p^{(1)} \bar{c}_{pq} S_q^{(1)}) dA \\ &= \int_A (F_j^{(0)} \bar{u}_j^{(0)} + F_j^{(1)} \bar{u}_j^{(1)}) dA \\ &+ \int_A (2b\rho b_j^{(0)} \bar{u}_j^{(0)} + \frac{2b^3}{3} \rho b_j^{(1)} \bar{u}_j^{(1)}) dA \\ &+ \int_C (p_j^{(0)} \bar{u}_j^{(0)} + p_j^{(1)} \bar{u}_j^{(1)}) ds \\ &- \int_A (2b\rho u_j^{(0)} \bar{u}_j^{(0)} + \frac{2b^3}{3} \rho u_j^{(1)} \bar{u}_j^{(1)}) dA \end{aligned} \quad (7)$$

where the edge tractions $p_j^{(0)}$ and $p_j^{(1)}$ are defined by

$$p_j^{(0)} = n_a T_{aj}^{(0)}, \quad p_j^{(1)} = n_a T_{aj}^{(1)}$$

and C is the intersection of the right cylindrical plate boundary with the middle plane of the plate and s the coordinate measured along the curve C .

For convenience in performing systematic numerical computations, the variational equation (7), which is often called the principle of virtual work, is expressed in the matrix form

$$\begin{aligned} & \int_A \{\bar{u}\}^T [M] \{u\} dA + \int_A \{\bar{S}\}^T [C] \{S\} dA \\ &= \int_A \{\bar{u}\}^T \{F\} dA + \int_A \{\bar{u}\}^T [M] \{b\} dA \\ &+ \int_C \{u\}^T \{p\} ds \end{aligned} \quad (8)$$

where $\{u\}$, $\{S\}$, $\{u\}$, $\{F\}$, $\{b\}$, and $\{p\}$ are the column vectors of stress, strain, displacement, face traction, body force and edge traction, respectively, as defined by

$$\begin{aligned} \{T\}^T &= \{T_p^{(0)}, T_p^{(1)}\}_{1 \times 11} \\ \{S\}^T &= \{S_p^{(0)}, S_p^{(1)}\}_{1 \times 11} \\ \{u\}^T &= \{u_j^{(0)}, u_j^{(1)}\}_{1 \times 6} \\ \{F\}^T &= \{F_j^{(0)}, F_j^{(1)}\}_{1 \times 6} \end{aligned}$$

$$\{b\}^T = \{b_j^{(0)}, b_j^{(1)}\}_{1 \times 6}$$

$$\{p\}^T = \{p_j^{(0)}, p_j^{(1)}\}_{1 \times 6}$$

and $[M]$ and $[C]$, the mass and stiffness matrices, are defined by

$$[M] = \begin{vmatrix} 2b\rho & 0 & 0 & 0 & 0 & 0 \\ 0 & 2b\rho & 0 & 0 & 0 & 0 \\ 0 & 0 & 2b\rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 2b^2\rho/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2b^3\rho/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2b^3\rho/3 \end{vmatrix}_{6 \times 6}$$

$$[C] = \begin{vmatrix} 2bc^{(0)} & 0 \\ 0 & \frac{2b^3}{3} c^{(1)} \end{vmatrix}_{11 \times 11}$$

$$[C^{(0)}] = [k_{(p)} k_{(q)} c_{pq}]_{6 \times 6}$$

$$[C^{(1)}] = [\bar{c}_{pq}]_{5 \times 5} \quad (9)$$

Isoparametric Finite Element Formulation

For finite element numerical analysis, the two-dimensional plate continuum is discretized into a number of small, but finite elements. And in each element, there are a number of designated points which are called nodal points.

The displacement field at any point in an element is called the element displacement and denoted by

$$\begin{aligned} \{u_p^h\}^T &= \{u_1^h = u_1^{(0)}, u_2^h = u_2^{(0)}, u_3^h = u_3^{(0)}, \\ &u_4^h = u_1^{(1)}, u_5^h = u_2^{(1)}, u_6^h = u_3^{(1)}\}. \end{aligned} \quad (10)$$

If the element nodal displacements are given, then the element displacement at any point within the element may be calculated approximately by interpolation

$$u_p(x_1, x_3, t) = \sum_{a=1}^{N_{en}} N_a(x_1, x_3) d_{pa}(t) \quad (11)$$

where $N_a(x_1, x_3)$ is a chosen function of x_1 and x_3 , and is called interpolation or shape function associated with the a th node, d_{pa} is the p th component of nodal displacement at a th node and N_{en} is the number of nodes per element.

For a four-node quadrilateral element, we let $N_{en} = 4$, and (11) becomes

$$u_p(x_1, x_3, t) = N_1^d p_1 + N_2^d p_2 + N_3^d p_3 + N_4^d p_4, \quad p = 1, 2, \dots, 6 \quad (12)$$

Displacement-nodal displacement relation

Substitution of (12) into (10) yields, for $N_{en} = 4$,

$$\{u^h\}_{6 \times 1} = [\phi]_{6 \times 24} \{d\}_{24 \times 1} \quad (13)$$

where

$$\{d\}^T = \{d_{11}, d_{12}, d_{13}, d_{14}, \dots, d_{61}, d_{62}, d_{63}, d_{64}\}_{1 \times 24}$$

$$[\phi] = \begin{bmatrix} N & 0 & 0 & 0 & 0 & 0 \\ 0 & N & 0 & 0 & 0 & 0 \\ 0 & 0 & N & 0 & 0 & 0 \\ 0 & 0 & 0 & N & 0 & 0 \\ 0 & 0 & 0 & 0 & N & 0 \\ 0 & 0 & 0 & 0 & 0 & N \end{bmatrix}_{6 \times 24}$$

$$N = \{N_1, N_2, N_3, N_4\}_{1 \times 4} \quad (14)$$

We see in (13) that the relationship between the element displacements at any point and the nodal displacements is established through the use of interpolation functions. This is the essence of the isoparametric finite element formulation, and by (13), we will, from now on, regard $\{d\}$, instead of $\{u_p^h\}$, as our unknown variables to be solved for.

Strain-nodal displacement relation

Substituting (1) into (3), and the result into (9)₂, we have the relationship between the element strains at any point and the nodal displacements, for $N_{en} = 4$,

$$\{S^h\}_{11 \times 1} = [B]_{11 \times 24} \{d\}_{24 \times 1} \quad (15)$$

where

$$[B] = \begin{bmatrix} N_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & N & 0 \\ 0 & 0 & N_3 & 0 & 0 & 0 \\ 0 & N_3 & 0 & 0 & 0 & N \\ N_3 & 0 & N_1 & 0 & 0 & 0 \\ 0 & N_1 & 0 & N & 0 & 0 \\ 0 & 0 & 0 & N_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & N_3 \\ 0 & 0 & 0 & 0 & N_3 & 0 \\ 0 & 0 & 0 & N_3 & 0 & N_1 \\ 0 & 0 & 0 & 0 & N_1 & 0 \end{bmatrix}_{11 \times 24}$$

$$N_1 = \{N_1, N_2, N_3, N_4\}_1$$

$$N_3 = \{N_1, N_2, N_3, N_4\}_3 \quad (16)$$

Nodal displacement equations of motion

By inserting (13) and (15) into (8), we obtain

$$\begin{aligned} \{\ddot{d}\}^T & \left(\int_{A_e} [\phi]^T [M] [\phi] dA \right) \{\ddot{d}\} \\ & + \{\ddot{d}\}^T \left(\int_{A_e} [B]^T [C] [B] dA \right) \{d\} \\ & = \{\ddot{d}\}^T \int_{A_e} [\phi]^T [F] dA + \{d\}^T \int_{A_e} \{\phi\}^T [M] \{b\} dA \\ & + \{d\}^T \int_{C_e} \{\phi\}^T \{p\} ds \end{aligned} \quad (17)$$

The above equation must hold for the arbitrary virtual nodal displacement $\{\ddot{d}\}$, therefore we have

$$[\bar{M}^e] \{\ddot{d}\} + [\bar{K}^e] \{d\} = \{\bar{F}^e\} \quad (18)$$

where the element mass matrix $[\bar{M}^e]$, element stiffness matrix $[\bar{K}^e]$, and element force vector $\{\bar{F}^e\}$ are defined by

$$\begin{aligned} [\bar{M}^e] &= \int_{A_e} [\phi]^T [M] [\phi] dA \\ [\bar{K}^e] &= \int_{A_e} [B]^T [C] [B] dA \\ \{\bar{F}^e\} &= \int_{A_e} [\phi]^T [F] dA + \int_{A_e} \{\phi\}^T [M] \{b\} dA \\ &+ \int_{C_e} \{\phi\}^T \{p\} ds \end{aligned} \quad (19)$$

Equation (18) will be employed to calculate initial stresses due to static loads or steady accelerations by setting $\{d\} = 0$ and to predict resonances and modes of free, incremental vibrations by setting $\{\bar{F}^e\} = 0$.

Initial Stresses

Fields of stress and displacement are computed by the finite element method in circular plates subject to three types of loading accommodated by (19):

(a) transverse loading $\{\bar{F}^e\} = \int_{A_e} [\phi]^T [F] dA$, (b) in-

plane body forces or in-plane accelerations

$\{\bar{F}^e\} = \int_{A_e} [\phi]^T [M] \{b\} dA$, (c) in-plane diametral

forces, $\{\bar{F}^e\} = \int_{C_e} \{\phi\}^T \{p\} ds$.

To establish the accuracy of present calculations, results are compared with those from analytical solutions and with experimental values.

Circular plate under transverse loading

Elastic and isotropic circular plate with two-point, three-point, and clamped edge supports are considered. A mesh with 192 elements for a circular plate is generated and shown in Fig. 1.

The plate is subject to either a concentrated force $P = 1$ lb. or a uniformly distributed load with intensity $W = P/A$.

The deflection at the center of the plate calculated by the present method is denoted by $u_2(\text{Num.})$ and that from Timoshenko's solution³ by $u_2(\text{Anal.})$. These deflections together with the geometrical and material properties of the plate are given in Fig. 2.

It may be seen that the comparison is close in all four cases. However, in the fourth row of the table in Fig. 2 where both the load distribution and edge support are uniform in spatial variation, the difference between the deflections from numerical and analytical solutions is the smallest.

Circular plate subject to in-plane body force

Stresses due to in-plane body force in an isotropic circular plate with four-point mount are studied.

For comparison, stress distributions along a diameter 45° from the x_1 axis while the body force is acting in the $-x_1$ direction are plotted in Fig. 3 and the corresponding ones from analytical solutions⁴ are shown in Fig. 4. It can be seen that agreement is good.

Circular plate subject to diametral forces

A circular plate of diameter d subject to a pair of diametral forces F whose direction is denoted by ψ , the azimuth angle measured from the x_1 axis, is shown in Fig. 5. In a previous study¹, an explicit formula for predicting changes in thickness-shear resonances due to diametral forces in rotated Y-cuts was derived

$$\frac{\Delta f}{f_0} = u_{1,1}^{(0)} + \frac{1}{2C_{66}} (C_{166}E_1^{(0)} + C_{266}E_2^{(0)} + C_{366}E_3^{(0)} + C_{466}E_4^{(0)}) \quad (20)$$

The above equation was used for calculating force sensitivity coefficient K_f as a function ψ for Y- and AT-cut quartz plates as shown in Figs. 6 and 7, respectively. In those calculations, stresses were approximated by the stresses obtained in an isotropic disk.

The effect of this approximation can be examined by calculating stresses at the center of the plate as functions of ψ for isotropic, Y- and AT-cuts circular disks. The results are shown in Figs. 8, 9, and 10 respectively. It may be seen that stresses in isotropic disk are very close to those in Y-cut except for values of $T_{11}^{(0)}$ near $\psi = 0$, but not as close to those in AT-cut.

We note that, in Y-cut, the value of C_{366} is much

greater than the rest of coefficients in (20). Therefore, the contribution to the frequency changes from $E_3^{(0)}$ or from $T_{33}^{(0)}$ is more dominant than the contributions from others. We see, indeed, the curve in Fig. 6 has the same shape of that of $T_{33}^{(0)}$ in Fig. 9.

For AT-cut, since C_{166} is much greater than other stiffness coefficients, $E_1^{(0)}$ or $T_{11}^{(0)}$ has the most contribution in (20). In this case, we see the experimental curve^{5,6} in Fig. 7 follows the $T_{11}^{(0)}$ -curve for AT-cut in Fig. 10 instead of the $T_{11}^{(0)}$ -curve for isotropic plate in Fig. 8.

Therefore we have demonstrated that the more accurate prediction in initial stresses provided by the finite element method improves the prediction of the changes of resonance frequencies due to initial stresses.

Free, Incremental Vibrations

By setting $\{\bar{F}^e\} = 0$, Equation (18) is employed to study the free vibrations of crystal resonators. Resonance frequencies and modes are computed for a square, AT-cut plate with mixed edge conditions and for a thin AT-cut strip with traction-free edge conditions. Corresponding two both cases, two- and one-dimensional close form solutions are obtained for the purpose of comparison.

Vibrations of a square, AT-cut plate

In order to obtain two-dimensional, close form solutions of (1) - (3) for checking the numerical results, the following mixed edge conditions are chosen.

$$\begin{aligned} t_{11}^{(0)} = u_2^{(0)} = u_3^{(0)} = t_{11}^{(1)} = u_2^{(1)} = u_3^{(1)} = 0 \\ \text{at } x_1 = \pm a \\ u_1^{(0)} = t_{32}^{(0)} = t_{33}^{(0)} = t_{31}^{(1)} = u_2^{(1)} = u_3^{(1)} = 0 \\ \text{at } x_3 = \pm c. \end{aligned} \quad (21)$$

To examine their respective merits, the four-node quadrilateral element (4Q) and eight-node quadrilateral element (8Q) are employed. Meshes consisting of sixty-four of four-node elements ($N=64$, 4Q) and sixteen of eight-node elements ($N=16$, 8Q) for a square plate are shown in Figs. 11 and 12, respectively.

Dimensionless frequencies $\Omega = \omega / (\frac{\pi}{2b} \sqrt{C_{66}/\rho})$ are computed by using three different mesh-element combinations: ($N=16$, 4Q), ($N=64$, 4Q) and ($N=16$, 8Q). These values are compared with the "exact" frequencies from the two-dimensional close form solutions and are listed in Table I for low frequencies and in Table II for high frequencies.

It is seen from both tables that values computed from ($N=16$, 8Q) have the best agreement.

Mode shapes corresponding to No. 684 listed in Table II are computed and compared in Fig. 13.

Vibrations of a thin, AT-cut strip

A mesh of sixty of eight-node elements is used to model a thin AT-cut quartz strip as shown in Fig. 14. For traction-free edge conditions, we require

$$T_{12}^{(0)} = T_{11}^{(1)} = 0 \quad \text{at } x_1 = \pm a \quad (22)$$

Frequencies calculated by using ($N=60$, $8Q$) are compared with the "exact" ones from the one-dimensional, close form solutions in Table III. It can be seen that the comparison is close, except some higher anharmonic overtones, like no. 8 and no. 10 in Table III, are missing from finite element calculations.

Mode shapes for no. 3 and no. 11 are plotted in Figs. 15 and 16, respectively. We see that comparison is close.

In conclusion, two computer programs have been developed based on a finite element method to predict initial stresses and to calculate resonances and modes of free, incremental vibrations in crystal resonators with triclinic crystal symmetry. The accuracy of the programs has been established by extensive comparison of computed results with those from analytical solutions and experimental values.

Acknowledgement

This work was supported by the U.S. Army Research Office, contract no. DAAG 29-84-0133.

References

1. P.C.Y. Lee, Y.S. Wang, and X. Markenscoff, "High-frequency vibrations of crystal plates under initial stresses", J. Acoust. Soc. Am., **57**, 95-105, (1975).
2. R.D. Mindlin, "An Introduction to the Mathematical Theory of Vibrations of Elastic Plates". Monogr. U. S. Army Signal Corps Eng. Labs., Fort Monmouth N.J. (1955).
3. S. Timoshenko and S. Woinowsky-Krieger, "Theory of Plates and Shells", McGraw-Hill Book Company, N.Y., p. 294, 1959.
4. P.C.Y. Lee, K.M. Wu and Y.S. Wang, "Effects of Acceleration on the Resonance Frequencies of Crystal Plates", J. Acoust. Soc. Am., **63** (4), 1039-1047, (1978).
5. A.D. Ballato, "Effects of Initial Stress on Quartz Plates Vibrating in Thickness Modes", Proc. 14th Am. Freq. Cont. Symp., pp 89-114, (1960).
6. C.R. Mingins, L. C. Barcus, and R.W. Derry, Rep. 1-20, Lowell Tech. Inst. Found. (1961-1966).

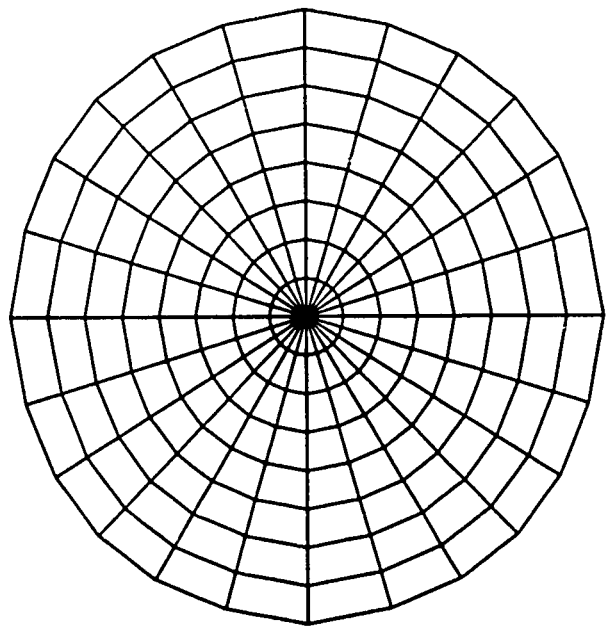


Figure 1: An undeformed mesh of 192 elements for a circular plate.

DISPLACEMENT AT CENTER OF CIRCULAR PLATE

SUPPORT	LOAD	ν	$t/0$	t/t	$10 \times T$	IS	HUNELL	u_z (Num.)	u_z (Anal.)	DIFF. (%)
2 pts.	P	0.25	1/100	1/6.25	1x1	1x1	192	0.03016*	0.02988*	0.9%
1 pts.	P	0.30	1/100	1/6.25	2x2	1x1	192	0.01630	0.01675	-2.7%
Clamped	P	0.25	1/100	1/6.25	2x2	1x1	156	0.005068	0.005124	-1.1%
Clamped	w	0.30	1/100	1/6.25	1x1	1x1	192	0.00975	0.00977	-0.2%

$$E = 1.092 \times 10^6 \text{ psi}, \quad R = 5 \text{ in.}, \quad P = 1.0 \text{ lb.}, \quad \text{MODP} = 6$$

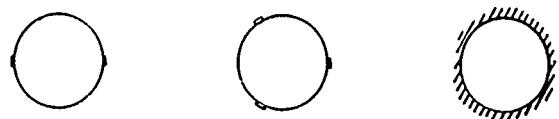


Figure 2: Comparison of central deflections of circular plate with various support conditions.

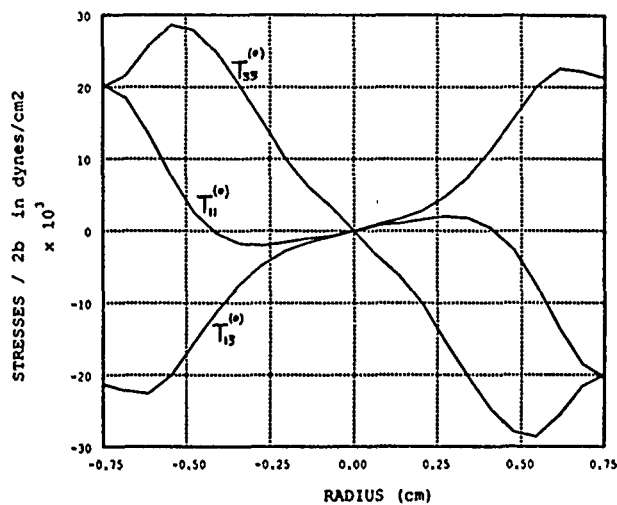


Figure 3: Stress distribution along a diameter 45° from the x_1 -axis of a circular plate subject to a body force in $-x_1$ direction. Numerical results.

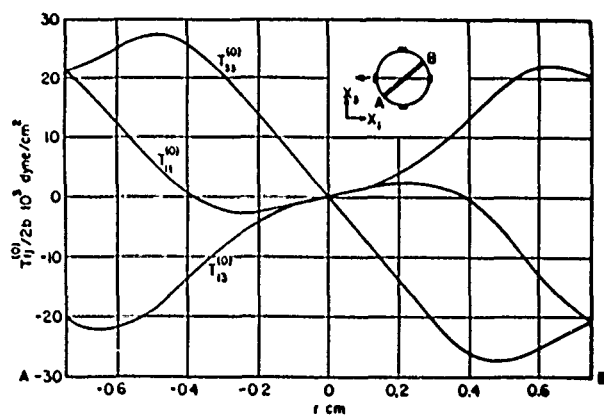


Figure 4: Stress distribution along a diameter 45° from the x_1 -axis of a circular plate subject to a body force in $-x_1$ direction. Analytical results.

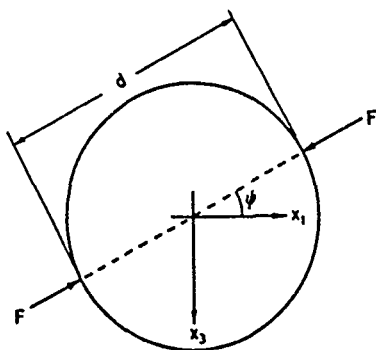


Figure 5: A circular plate under diametral forces.

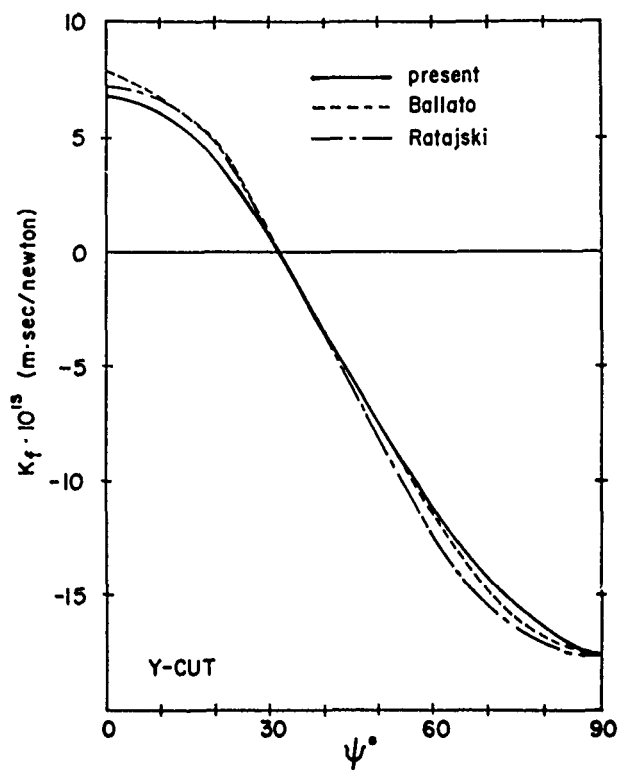


Figure 6: Force sensitivity coefficient K_f as a function of the azimuth angle ψ for Y-cut plate.

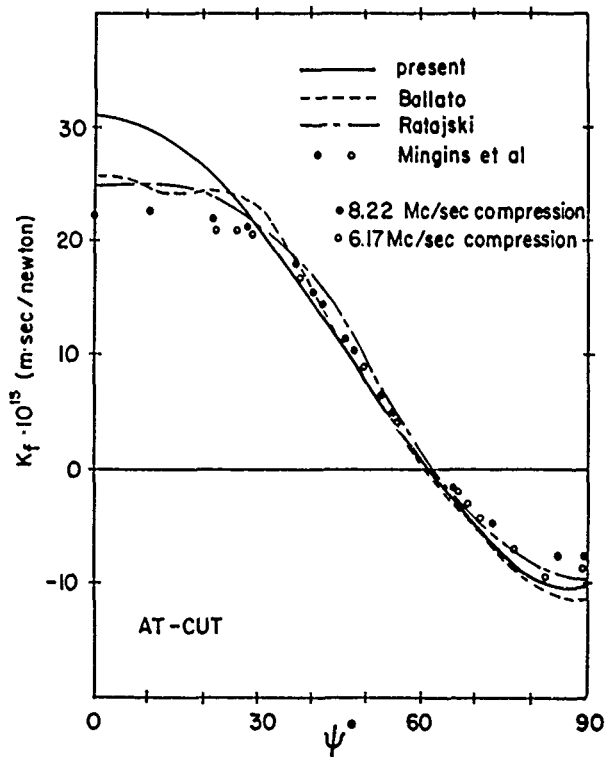


Figure 7: K_f vs ψ for AT-cut plate.

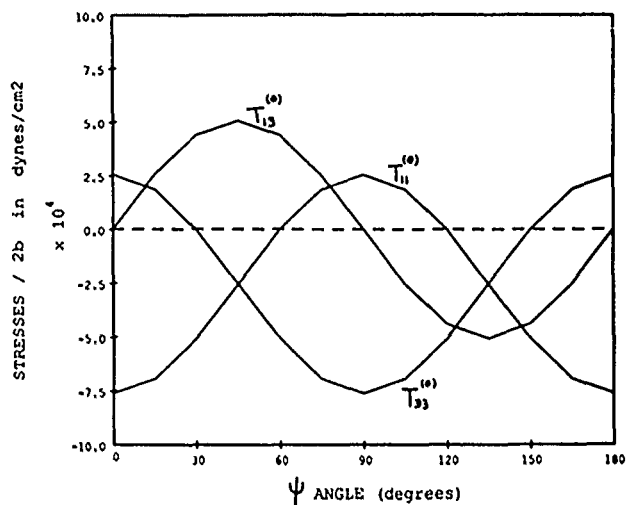


Figure 8: Stresses at center of plate as a function of azimuth angle ψ for isotropic plate.

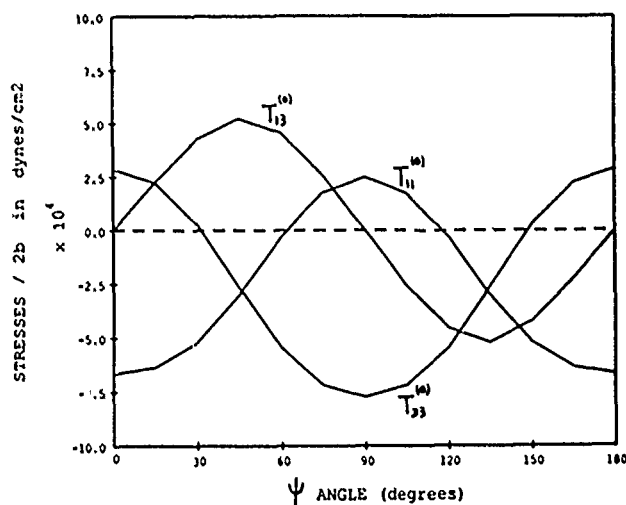


Figure 9: $T_{ab}^{(0)}$ vs ψ for Y-cut plate, $a, b = 1, 3$.

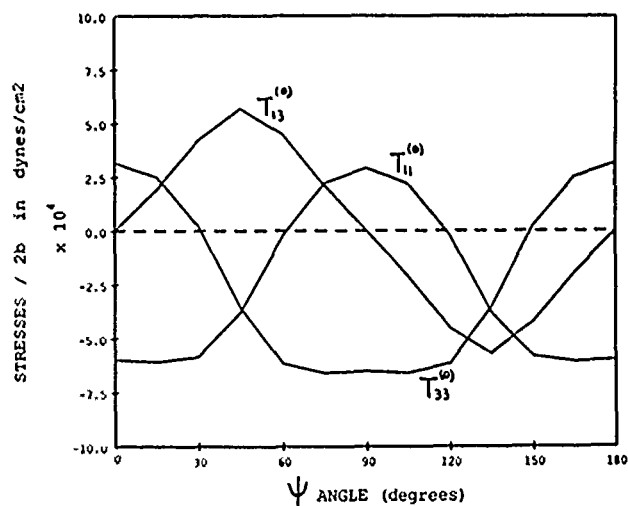


Figure 10: $T_{ab}^{(0)}$ vs ψ for AT-cut plate, $a, b = 1, 3$.

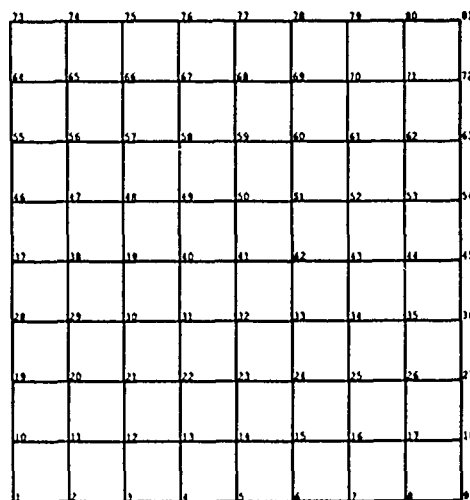


Figure 11: An undeformed square mesh of sixty-four of four-node elements.

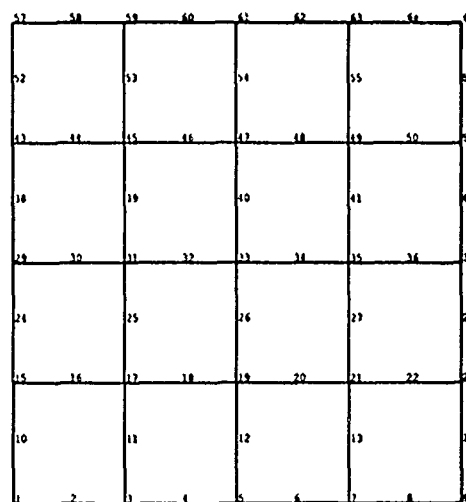


Figure 12: An undeformed square mesh of sixteen of eight-node elements.

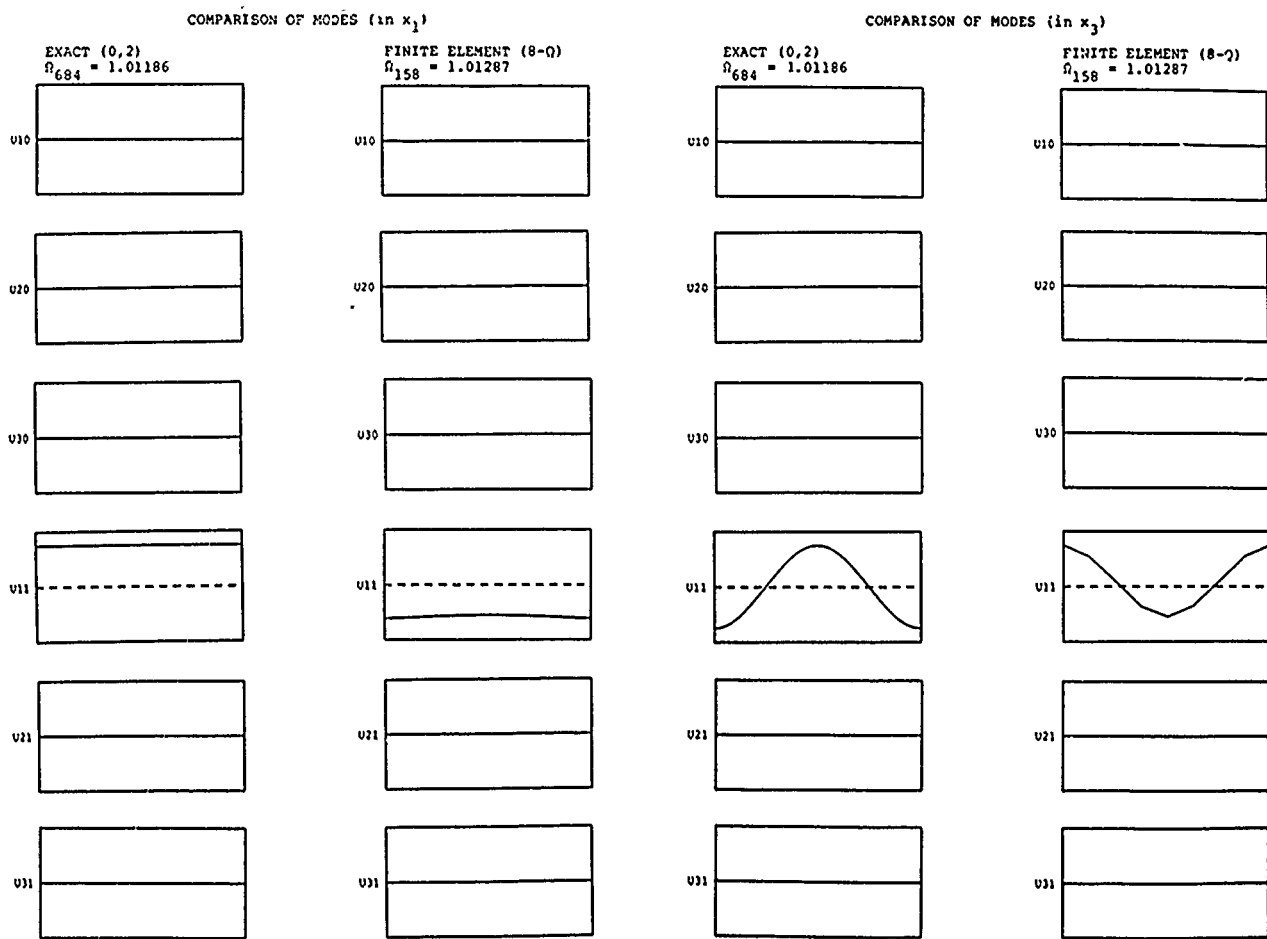


Figure 13: Comparison of modes along the x_1 and x_3 axes in a square AT-cut at frequency no. 684 listed in Table II.

Table I

AT-CUT SQUARE PLATE
Mixed Boundary Conditions
Low Frequencies Comparison Between Exact
and finite element solutions

No	MODE		EXACT $\Omega(x10^{-2})$	FINITE ELEMENT		
				4Q, N=16 $l/t=5.0$ $\Omega(x10^{-2})$	4Q, N=64 $l/t=2.5$ $\Omega(x10^{-2})$	8Q, N=16 $l/t=5.0$ $\Omega(x10^{-2})$
	x_1	x_3				
1	1	0	0.39082	1.14124	0.65093	0.39556
2	1	1	0.95014	1.88910	1.23768	0.95499
3	2	0	1.54169	4.96562	2.66002	1.63409
4	2	1	2.17477	5.12322	3.11806	2.25409
5	1	2	2.27234	6.24984	3.42045	2.36434
6	3	0	3.39299		6.18417	3.87629
7	2	2	3.67504	8.06247	4.93139	3.90422
8	3	1	4.03063	9.97358	6.52822	4.50214
9	1	3	4.30040		7.39728	4.83560
10	3	2	5.63597		7.97413	

Table II

AT-CUT SQUARE PLATE
Mixed Boundary Conditions
High Frequencies Comparisons between Exact
and Finite Element Solutions

No	MODE		EXACT Ω	FINITE ELEMENT		
				4Q, N=16 $l/t=5.0$ Ω	4Q, N=64 $l/t=2.5$ Ω	8Q, N=16 $l/t=5.0$ Ω
	x_1	x_3				
666	0	0	1.00000	1.00086	1.00075	1.00042
672	0	1	1.00298	1.00475	1.00453	1.00388
675	1	0	1.00476	1.00666	1.00620	1.00556
678	1	1	1.00753	1.01148	1.01083	1.00940
684	0	2	1.01186	1.01543	1.01366	1.01287
685	1	2	1.01585	1.02166	1.01989	1.01822
688	2	0	1.01883	1.02324	1.02107	1.01973
690	2	1	1.02098	1.02691	1.02446	1.02257
692	0	3	1.02649	1.03861	1.03148	1.02866
694	2	2	1.02763	1.03604	1.03290	1.03028

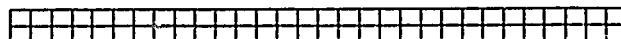


Figure 14: An undeformed mesh of sixty elements for a thin strip.

Table III

AT-CUT STRIP IN x_1 -DIRECTION
 Free Boundary Conditions at $x_1 = \pm a$
 High Frequencies Comparison between Exact
 and Finite Element Solution.

$$a/b = 20, \quad c/b = 1$$

$$N = 60, \quad (30 \text{ in } x_1, 2 \text{ in } x_3)$$

No	MODE TYPE	EXACT Ω	FEM(8-Q) Ω
1	Sym	1.00290	
2	Asym	1.00461	1.01197
3	Sym	1.00476	1.00406
4	Asym	1.02414	1.03299
5	Sym	1.03347	1.03547
6	Sym	1.06170	1.07353
7	Asym	1.06642	1.06933
8	Asym	1.07999	
9	Asym	1.10511	1.11567
10	Sym	1.10625	
11	Sym	1.15046	1.16044

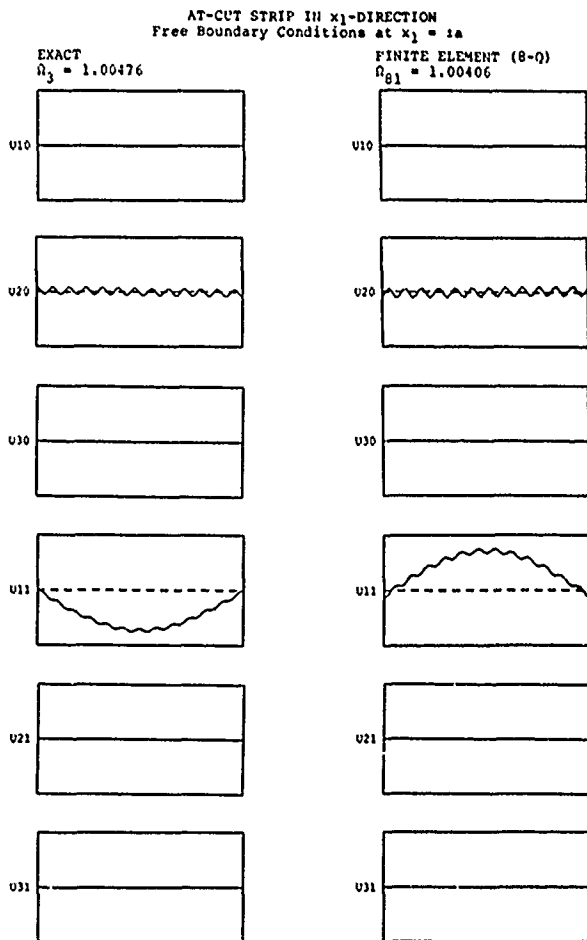


Figure 15: Comparison of modes in a thin, AT-cut strip at frequency no. 3 listed in Table III.

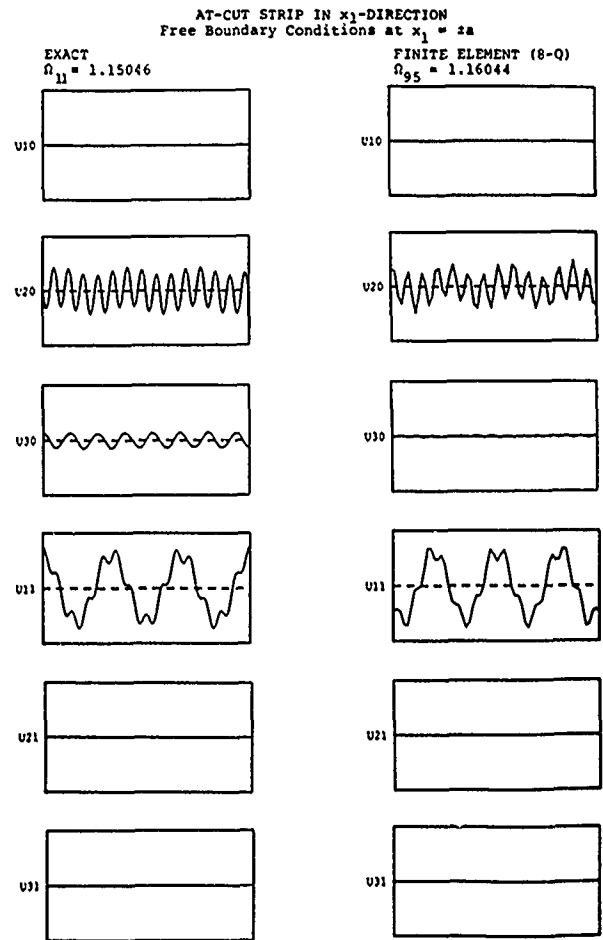


Figure 16: Comparison of modes in a thin, AT-cut strip at frequency no. 11 listed in Table III.

A TRANSMISSION LINE MATRIX MODEL FOR AT QUARTZ THICKNESS SHEAR DEVICES

Lawrence N. Dworsky

Motorola, Inc.
Schaumburg, Illinois

Summary

A computer program has been written for simulating arbitrary monolithic crystal filter structures on AT quartz. The program assumes the exclusive existence of X thickness-shear vibrations and a uniform quartz plate. The surface of the plate is subdivided into small rectangular regions over each of which electrode thickness is considered uniform. In each region it is therefore possible to separate the thickness dependent term of the displacement function of the governing wave equation and reduce the wave equation to a two dimensional equation. The 2-dimensional space described by this equation is then approximated by a pair of intersecting acoustical transmission lines. The electric circuit is coupled into the cell at the intersection of these lines.

The program has proven to be accurate in simulating the passband and predominant spurious mode responses of various monolithic filter designs. Examples of simulated and actual responses are given.

Introduction

For AT quartz thickness shear mode devices, Tiersten has shown [ref 1] that the wave equation describing the displacement is

$$\kappa_n \frac{\partial^2 u_x}{\partial x^2} + c_{55} \frac{\partial^2 u_x}{\partial z^2} + \bar{c}_{66} \frac{\partial^2 u_x}{\partial y^2} + \rho \omega^2 u_x = 0 \quad [1]$$

where all terms not defined follow reference [1]. From this same reference, we see that in a region of uniform thickness, we may write u as a product of $U(x,z)$ and a thickness dependent term which approximately satisfies the free-surface boundary conditions

$$u_x(x,y,z) = U(x,z) \sin\left(\frac{n\pi y}{2h}\right) \quad [2]$$

Substituting [2] into [1], we get a reduced wave equation,

$$\kappa_n \frac{\partial^2 U}{\partial x^2} + c_{55} \frac{\partial^2 U}{\partial z^2} + \left[\rho \omega^2 - \bar{c}_{66} \left(\frac{n\pi}{2h} \right)^2 \right] U = 0 \quad [3]$$

Let us temporarily consider cases where all relevant dimensions in X are much larger than those in Z , so that partial derivatives with respect to X are much smaller than those with respect to Z , and may be ignored for the moment:

$$c_{55} \frac{d^2 U}{dz^2} + K_y^2 U = 0 \quad [4]$$

where

$$K_y^2 = \rho \omega^2 - \bar{c}_{66} \left(\frac{n\pi}{2h} \right)^2 \quad [5]$$

Equation [4] is a one-dimensional wave equation, more commonly known as the "transmission line" equation. Furthermore, if we note that the X - Z shear stress, T , is given as

$$T = c_{55} \frac{dU}{dz} \quad [6]$$

then [4] may be broken into the two equations [6] and

$$\frac{\partial T}{\partial z} + K_y^2 U = 0 \quad [7]$$

The above may be considered equivalent to an electrical transmission line with L and C being the inductance and capacitance per unit length, respectively. For such a line,

$$-j\omega L = \frac{1}{\omega C} \frac{\partial V}{\partial z} \quad [8]$$

and

$$\frac{\partial}{\partial z} (-j\omega L) + \omega C V = 0 \quad [9]$$

are the equations analogous to [6] and [7].

The analogy above is functionally equivalent to the one originally proposed by Mason [ref 2], but in this case the electrical voltage corresponds to the mechanical displacement. It was used by Dworsky [ref 3] to model one (surface) dimensional crystal devices. In principal, an arbitrary device is approximated by connecting together many sections of transmission line, each of which represents a region of uniform parameters on the crystal plate. Of particular relevance is the fact that the resonances of structures modeled using the 1 dimensional transmission line analogy are exactly the resonances which would be found if the differential equations were solved directly, using continuous stress and displacement boundary conditions at the interfaces of the sections.

The intent of the model to be discussed is the treatment of the more general, 2 (surface) dimensional problem. In order to do this, we wish to generalize the transmission line analogy to that of sections of transmission line interconnected in a rectangular grid. This is known as a (2 dimensional) transmission line matrix (TLM).

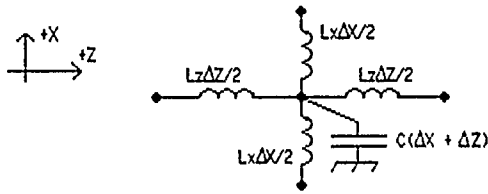


Figure 1. TLM "Cell".

Consider a rectangular grid of inductors and capacitors, as shown in Figure 1. The inductance per unit length in the X direction is L_x while that in the Z direction is L_z . The capacitance at the junction of the two lines is $2C$. The circuit equations which describe Figure 1 are

$$-jI_x = \frac{1}{\omega L_x} \frac{\partial V}{\partial x} \quad [10]$$

$$-jI_z = \frac{1}{\omega L_z} \frac{\partial V}{\partial z} \quad [11]$$

and

$$\frac{\partial I_x}{\partial x} + \frac{\partial I_z}{\partial z} = -j\omega 2CV \quad [12]$$

Combining the above 3 equations, we get the 2 dimensional wave equation

$$\frac{1}{\omega L_x} \frac{\partial^2 V}{\partial x^2} + \frac{1}{\omega L_z} \frac{\partial^2 V}{\partial z^2} + 2\omega CV = 0 \quad [13]$$

Equation [13] shows that, in some sense, the 2 dimensional TLM grid approximates a 2-dimensional space. This model has been used for generating approximate solutions to Maxwell's equations [ref 4]. Comparing [13] to [3], we see that we have an analogy to the reduced 3 dimensional crystal wave equation. By direct comparison of these equations,

$$L_x = \frac{1}{\omega M_n}, \quad L_z = \frac{1}{\omega C_{33}}, \quad C = \frac{\rho\omega^2 - \bar{C}_{66}k_y^2}{2\omega} \quad [14]$$

The characteristic impedances of the 2 lines are

$$Z_{0x} = \left(\frac{L_x}{C} \right)^{1/2}, \quad Z_{0z} = \left(\frac{L_z}{C} \right)^{1/2} \quad [15]$$

and the propagation constants are

$$\gamma_x = \omega(jL_x C)^{1/2}, \quad \gamma_z = \omega(jL_z C)^{1/2} \quad [16]$$

It will be shown below that the propagation velocities are a function of, among other things, the electrical lengths of the lines. It is therefore desirable to keep the X and Z propagation times between grid points identical. For S_x the X direction grid spacing and S_z the Z direction spacing, using [14] and [16],

$$S_x = S_z \left(M_n / C_{33} \right)^{1/2} \quad [17]$$

Let us define a "cell" as a rectangular region on the surface of the crystal plate of dimensions (S_x, S_z). The electrical circuit analog to the TLM approximation of the wave equation in this cell is represented by the same circuit as Figure 1. The circuit values are given by [14]. The value of the voltage at the center of the cell, i.e. at the junction of the 5 components, will represent the displacement, U , for the cell. Note that the parameter h will take into account both the thickness of the quartz and the thickness of any deposited mass loading. Materials other than quartz may be treated by scaling their thicknesses to the ratio of their mass density to that of quartz. The value of h in a given cell is considered constant. A device design is approximated by building up a rectangular array of cells, each with its own parameters.

Before considering electrical connections to the acoustical analog, boundary conditions, etc., it is important to look at the propagation properties of the TLM structure. Remember that in the 1 dimensional transmission line structure (at a given frequency for the acoustical analog) the propagation velocity is a constant. This is why the 1 dimensional structure predicts correct resonances.

Consider a 2 dimensional TLM grid as defined above. For a wave propagating in (say) the X direction the wave front will reach a row of node points, and from each node launch scattered wave fronts in all 4 directions. By symmetry, then, each node will see an incident wave from the + and - Z directions identical to the wave that same node would see if we had a 1 dimensional structure with a pair of lines $S_z/2$ long, open-circuited, attached to the node. For any cell, therefore, we may write the chain (ABCD) network parameters as

$$\begin{aligned} & \begin{bmatrix} \cos(\theta/2) & j\sin(\theta/2) \\ j\sin(\theta/2) & \cos(\theta/2) \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 2j\tan(\theta/2) & 1 \end{bmatrix} \\ & \times \begin{bmatrix} \cos(\theta/2) & j\sin(\theta/2) \\ j\sin(\theta/2) & \cos(\theta/2) \end{bmatrix} \quad [18] \\ & = \begin{bmatrix} [\cos^2(\theta/2) - 3\sin^2(\theta/2)] & - \\ - & [\cos^2(\theta/2) - 3\sin^2(\theta/2)] \end{bmatrix} \end{aligned}$$

The ABCD parameters used above are defined in the usual manner,

$$\begin{bmatrix} V_1 \\ 1_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ -1_2 \end{bmatrix} \quad [19]$$

For a periodic structure, $V_1/i_1 = V_2/-i_2$. Substituting this relation back into [19], and letting $r = V_1/V_2$, we get

$$r^2 - (A + D)r + 1 = 0 \quad [20]$$

Substituting [18] into [20] and solving the quadratic equation,

$$r = \left[\cos(\theta) - 2\sin^2(\theta/2) \right] + j \left[1 - \left[\cos(\theta) - 2\sin^2(\theta/2) \right]^2 \right]^{1/2} \quad [21]$$

The phase angle of r describes the propagation time between the grid points:

$$\beta = \tan^{-1} \left(\frac{\text{Imag}(r)}{\text{Real}(r)} \right) = \frac{\omega S_z}{V_{eq}} \quad [22]$$

If we relate this to the electrical length of the line connecting the grid points,

$$\theta = \omega S_z / V_0 \quad [23]$$

where V_0 is the propagation velocity on this line, we get

$$\frac{V_{eq}}{V_0} = \frac{\theta}{\beta} \quad [24]$$

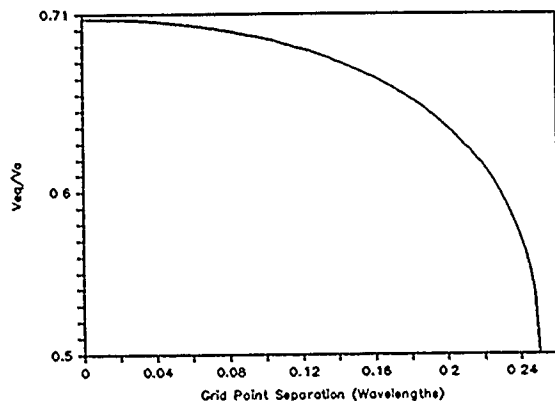


Figure 2. Relative Wave Velocity vs Grid Spacing

Figure 2 shows the ratio of propagation velocity along the grid to propagation velocity of the line versus the electrical length of the line in fractions of a wavelength. As may be seen, for very dense grids this ratio approaches .707, whereas as the separation

approaches $1/4$ wavelength, the ratio falls off. For a separation $> 1/4$ wavelength, there is a cutoff region where no propagation occurs.

In terms of practical AT quartz devices, the electrical line length is itself a function of frequency. However, since electrode size is not the dominant factor in determining device resonance, small variations in effective velocity (i.e. effective electrode size) don't seriously perturb the results. The most significant results of this calculation are 1: grid separation must be small compared to $1/4$ wavelength on the lines, and 2: an electrode must be sized .707 times its actual physical length for correct results.

The appendix shows an example of the magnitude of error produced by the change in equivalent velocity with frequency.

II. The Full Model

Consider again the cell model of Figure 1. Assume that there are an infinite number of identical cells interconnected, covering the entire $X - Z$ plane - i.e. an "infinite plate" of AT quartz. Noting that the electrical lengths of the lines in both the X and the Z directions are the same, we may drop the subscripts on the propagation constant and the line length. The admittance seen "looking into" the node at the center of any one cell is simply

$$Y_{in} = 4Y_0 \tanh(\gamma S/2) \quad [25]$$

At resonance this admittance must be 0 (the displacement goes to infinity). Substituting the values developed in the previous section,

$$\frac{j4C_{33}k}{\sqrt{2}} \tanh \left(\frac{kS}{2\sqrt{2}} \right) \quad [26]$$

Equation [26] is satisfied if $K_y = 0$, i.e.

$$\omega^2 = \omega_D^2 = \bar{C}_{66} \left(\frac{n\pi}{2h} \right)^2 \quad [27]$$

Also,

$$\frac{dY_{in}}{d\omega} = \frac{j4C_{33}}{\sqrt{2}} \frac{\omega\rho}{C_{33}k} \left[\tanh \left(\frac{kS}{2\sqrt{2}} \right) + \frac{kS}{2\sqrt{2}} \text{sech}^2 \left(\frac{kS}{2\sqrt{2}} \right) \right] \quad [28]$$

which at resonance becomes

$$\left. \frac{dY_{in}}{d\omega} \right|_{k=0} = j2\rho\omega_S \quad [29]$$

If we wish to extend the model so as to make the node at the center of the cell approximate the electrical input to the cell, we note that the resonance of the infinite plate looks like a series resonance electrically.

If the "motional inductance" of this series resonator is L_m , then

$$Z \Big|_{k=0} = 0, \quad \frac{dZ}{d\omega} \Big|_{k=0} = j2L_m \quad [30]$$

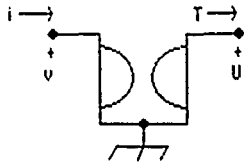


Figure 3. Gyration Circuit Conventions

The circuit element which will transfer the admittance properties described above to the desired electrical properties is the gyrator (Figure 3), with properties

$$U = gI, \quad T = v/g \quad [31]$$

From [31],

$$Z = \frac{v}{i} = g^2 \frac{T}{U} = g^2 Y_{in} \quad [32]$$

and

$$\frac{dZ}{d\omega} = g^2 \frac{dY_{in}}{d\omega} \quad [33]$$

Combining [33] and [29], we find that

$$g = \left[\frac{L_m}{\rho \omega s} \right]^{1/2} \quad [34]$$

Putting aside the question of boundary conditions for a moment, we now have all the information necessary to "assemble" an analog circuit model of an arbitrary structure:

1. Divide the structure into rectangular regions (cells), noting that the total thickness (quartz + electrode) must be constant in any given region. Calculate the various parameters (k_y , Z_0 , etc.) in each of these regions.

2. "Assemble" a TLM grid made up of the individual cells, forming a rectangular pattern.

3. Write a set of nodal circuit equations describing the TLM grid. (Since Y has been used above, let us use A as the nodal matrix). This nodal set will be of the form

$$[A_{ij}] [U_j] = [T_i] \quad [35]$$

where the U_i are the unknown displacements, and the form of the A_{ij} will be given below. The forcing vector, T , is 0 everywhere except in the cells where an electrical source is connected. In these latter cells, we

apply a source of 1 volt. Using [31], this means that we set the chosen $T = 1/g$.

4. Solve the set of equations. This is not necessarily a trivial problem, and will be discussed further below.

5. Since the U_i are now known, we may calculate the total electric current in the "source cells" (i.e. the chosen T_i cells above) as

$$I_1 = \sum_k \frac{U_k}{g_k} \quad [36]$$

where the summation, k , is over the source cells. Since the input voltage was originally set to 1, this current is the input admittance of the structure.

6. If we are dealing with a 2-port network, the current found above is identically Y_{11} . Repeating the calculation [36] over the "output cells" (i.e. the cells which form the electrode area connected to the output of the network), we get Y_{21} . By reciprocity, $Y_{21} = Y_{12}$. If our structure is electrically symmetric, $Y_{22} = Y_{11}$. If not, we must go back and reverse the definitions of source and output cells, and repeat the entire calculation. In any case, we now have the Y parameters of the arbitrary structure as seen from the electrical ports, and may calculate the response in any electrical circuit environment.

The form of the A_{ij} terms is found from a straightforward application of circuit theory. A_{ij} is a nodal circuit matrix. This means that each diagonal term, A_{ii} , is equal to the sum of the diagonal nodal matrix terms of all 2-port devices connected to node i and all 1 port devices connected from node i to ground, while each off-diagonal term, $A_{ij} = A_{ji}$, is equal to the sum of the off-diagonal terms of all 2-port devices connected between nodes i and j .

It turns out to be most convenient to work in ABCD parameters, so we note that

$$\begin{bmatrix} Y_{1j} \end{bmatrix} = \begin{bmatrix} D/B & -1/B \\ -1/B & A/B \end{bmatrix} \quad [37]$$

If we "follow" a line from any one node to any adjacent node, we see that we cross a cell boundary mid-way. This line is therefore made up of two sections, possibly different from each other. Designating side i as the "left" and side j as the "right", the ABCD parameters for the line are

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \cosh(\gamma_1 S/2) & Z_{01} \sinh(\gamma_1 S/2) \\ Y_{01} \sinh(\gamma_1 S/2) & \cosh(\gamma_1 S/2) \end{bmatrix} \times \begin{bmatrix} \text{same for } j \end{bmatrix} \quad [38]$$

from which we get

$$B = Z_{0j} \sinh\left(\frac{\gamma_j S}{2}\right) \cosh\left(\frac{\gamma_i S}{2}\right) + Z_{0i} \sinh\left(\frac{\gamma_i S}{2}\right) \cosh\left(\frac{\gamma_j S}{2}\right) \quad [39]$$

and

$$D = Y_{01} Z_{0j} \sinh\left(\frac{\gamma_1 S}{2}\right) \sinh\left(\frac{\gamma_1 S}{2}\right) + \cosh\left(\frac{\gamma_1 S}{2}\right) \cosh\left(\frac{\gamma_1 S}{2}\right) \quad [40]$$

Note that if the two adjacent cells are identical, the above simplifies to

$$B = Z_0 \sinh(\gamma S), \quad D = \cosh(\gamma S) \quad [41]$$

There is never a need to explicitly substitute [39] and [40], or [41], into [37]. The computer program can simply calculate B and D and then substitute the results into [37].

At the outer edges of the TLM grid there must be some type of boundary conditions. Grounding the outer edges of all cells would correspond to a plate which is clamped on all edges - not a physically realizable condition. In general we would like to simulate an infinite plate. The infinite plate assumption is usually accurate for overtone devices, but of uncertain accuracy on fundamental mode devices.

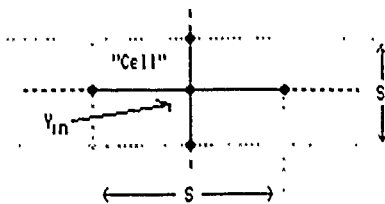


Figure 4. Infinite Plate of TLM Cells

Consider, as shown in Figure 4, an infinite region of identical cells. The impedance seen looking away from any one grid point in any direction is Z_t . As may be seen, Z_t must be equal to the input impedance of a length S of uniform line, terminated in the impedance $Z_t/3$:

$$Z_T = Z_0 \left[\frac{\frac{Z_T}{3} + Z_0 \tanh(\gamma S)}{Z_0 + \frac{Z_T}{3} \tanh(\gamma S)} \right] \quad [42]$$

from which

$$Z_T = Z_0 \left(-\coth(\gamma S) + \sqrt{\coth^2(\gamma S) + 3} \right) \quad [43]$$

Unfortunately, Z_t does not correctly represent the infinite plate condition for a real device. In the general case of multi-resonator structures, there are waves propagating in all directions near a discontinuity - e.g. an electrode edge. It has been found empirically that if a structure is surrounded by at least three rows of cells which are in turn terminated by Z_t on their outward-facing edges, a reasonable infinite plate approximation is realized.

The last item to be considered is that of losses. There are no rigorous formulae for the losses of AT

quartz, especially when considering the damping of electrode metalization. Based upon experience, it is possible to come up with some numbers for the quartz and metal losses, and combine them in a (hopefully) meaningful way. Such a formula is usable only for quartz and metal thicknesses near those for which the formula was developed, and for devices processed in a manner identical to those for which the formula was developed. With these caveats in mind, we propose the formula

$$Q = \frac{10^{10}}{f} \left[\frac{h_b + h_L}{h_b/1500 + h_L/6.5} \right] \quad [44]$$

The finite Q is introduced into the equations by making the C_{ij} complex. For lack of better information, we let

$$C_{ij} \rightarrow C_{ij} (1 + j/Q) \quad [45]$$

for all C_{ij} .

III. Programming Considerations

The principal chores of the computer program to implement this model are 1) to collect an input geometry description, 2) to convert this description to the matrix equation [35] and 3) to solve this equation. Step 1 is a tedious but nonetheless straightforward chore which presents no real challenge to the programmer.

Step 2 requires some thought before recklessly charging ahead. Since all variables involved are in general complex, we must use complex variable types. Also, since we will casually ask for frequency resolution greater than 1 part in 10^7 , the intermediate calculations must be held to much better resolution, i.e. we need double precision numbers. In Fortran, this means that we will tie up 16 bytes of storage for each (Complex*16) number. Now, consider a practical case. We could easily need a 40×50 grid, i.e. 2000 grid points, to describe what we want. This means that we must solve a simultaneous equation problem having 2000 variables. The coefficient matrix, in principal, has 4 million variables - which translates into 64 million bytes of storage! Clearly this is not tractable. We are saved, however, by the sparsity of the problem. In each equation, there are only 5 non-zero terms: the diagonal (self) term, and the 4 off-diagonal terms describing interconnections to the 4 adjacent nodes. This means that in our 2000 variable problem, there are only $5 \times 2000 = 10000$ non-zero numbers, or 160,000 bytes of storage required. Furthermore, if this number is a problem, we may note that there are not 10000 different numbers. If we consider a situation with a crystal plate and a single metalization on each side (resulting in regions of top and bottom metalization, and also regions of only one sided metalization) then there are only 16 different non-zero numbers to consider. A simple 1 byte integer for each of the 10000 non-zero coefficients pointing to a 16 number table reduces the coefficient matrix storage requirement down to essentially 10,000 bytes. Finally, we note that A_{ij} must be symmetric, so there are only 3 stored numbers per equation required.

Step 3, solving the system of equations, requires a scheme which does not destroy the sparseness of the

coefficient matrix. Unfortunately, Gauss-Seidel iterations are in general not convergent in this application. The method chosen was a gradient-iteration technique referred to as the "conjugant gradient" technique [ref 5]. This technique, in addition to not modifying the coefficient matrix in any way, also has the property of converging towards the correct solution with successive iterations - theoretically reaching the exact solution after N iterations (N being the order of the problem). This means that in many cases a satisfactory solution is reached after less than N iterations. It should be noted that an exhaustive study was not made of this issue, and there are very many alternate approaches in the literature.

When setting up multi-pole monolithic filter problems with more than one gap size, it is usually necessary to set S_z (or S_x) to be the largest common divisor (approximately) of all of the gaps. This can lead to very large, slow simulations. Now, while the exact filter layout is necessary for obtaining the correct passband, this program is not necessary for modelling the passband - the passband usually performs as designed. The principal need for this program is the spurious mode response pattern. It has been found empirically that the spurious mode response can be approximated very well by simply replacing all gaps by an average value, and then setting S_z (or S_x) to this value. In this way the number of cells required is kept to a minimum, and program run time is optimized.

Actual Results

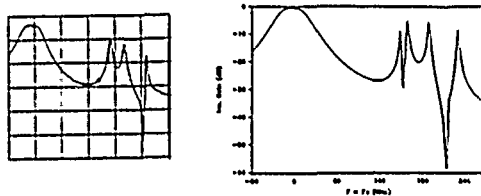


Figure 5. 2-Pole, N = 3, Monolithic Filter
(Left: Actual Data Right: Simulation)

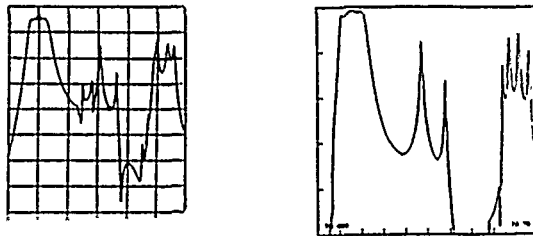


Figure 6. 4-Pole, N = 3, Monolithic Filter
(Left: Actual Data Right: Simulation)

Figures 5 and 6 show two comparisons of actual experimental device data to computer simulation results. In the first case, a 2-pole monolithic filter, the agreement is excellent. In the second case, a 4-pole monolithic filter, the passband and predominant spurious responses agree very well. The actual devices exhibited some small close-in responses which the simulation did not predict. This discrepancy has not yet been resolved.

IV. Comments

A transmission line matrix model has been described which may be used for simulating the response of arbitrary AT thickness shear resonators and/or monolithic filter structures. Some aspects of writing a computer program to implement the model have been discussed, and alternatives for minimizing computer memory requirements presented. The model has proven to be accurate for overtone mode structures, where energy trapping effectively makes the quartz plate edges appear infinitely far away.

V. References

1. H. F. Tiersten, "Analysis of Trapped Energy Resonators Operating in Overtones of Coupled Thickness-Shear and Thickness-Twist." J. Acoust. Soc. Am., Vol 59, P. 879, 1976.
2. W. P. Mason, "Equivalent Electromechanical Representation of Trapped Energy Resonators." Proc. of the IEEE, Vol 57, #10, Oct. 1969.
3. L. N. Dworsky, "Discrete Element Modeling of AT Quartz Devices." Proc. 32nd Symposium on Frequency Control, U.S. Army Electronics Command, Fort Monmouth, NJ, P. 142, 1978.
4. P. B. Johns & R. L. Beurle, "Numerical Solution of 2 Dimensional Scattering Problems Using a Transmission Line." Proc. of the IEE, Vol 118, Nr. 9, Sept. 1971.
5. A. Jennings, "Matrix Computation for Engineers and Scientists." John Wiley & Sons, New York, 1977.

Appendix: Grid Size Error Estimate

Consider equation [13], a 2 dimensional wave equation:

$$\frac{1}{L_x} \frac{\partial^2 v}{\partial x^2} + \frac{1}{L_z} \frac{\partial^2 v}{\partial z^2} + \omega^2 C v = 0 \quad [A.1]$$

Remembering that the above is the steady state form of a wave equation with an assumed sinusoidal time dependence, we can in general write its solution as

$$v = \exp \{ j(\omega \sqrt{L_x C} x + \omega \sqrt{L_z C} z - \omega t) \} \quad [A.2]$$

For a wave travelling in, say, the Z direction we may calculate a velocity by setting the phase of the above equal to a constant

$$\omega(\sqrt{L_x C} x + \sqrt{L_z C} z - t) = \text{const} \quad [A.3]$$

from which

$$v_z = \frac{\partial z}{\partial t} = 1/\sqrt{L_z C} \quad [A.4]$$

In this case the velocity is strictly a function of the material parameters of the space.

Now, let us consider equation [3], where p is the frequency of an infinite plate whose thickness is equal to that of the propagating region under study:

$$M_n \frac{\partial^2 U}{\partial x^2} + C_{33} \frac{\partial^2 U}{\partial z^2} + \rho(\omega^2 - p^2)U = 0 \quad [A.5]$$

with

$$\omega > p \quad [A.6]$$

Directly parroting [A.2] - [A.4],

$$U = \exp \left\{ j \left(\frac{\rho(\omega^2 - p^2)}{C_{33}} \right)^{1/2} x + j \left(\frac{\rho(\omega^2 - p^2)}{C_{33}} \right)^{1/2} z - j\omega t \right\} \quad [A.7]$$

$$v_z = \sqrt{C_{33}/\rho} \frac{1}{\sqrt{1 - (p/\omega)^2}} \text{ M/Sec} \quad [A.8]$$

As a sample calculation, consider a 50 MHz resonator. For $p \ll f$, the wavelength is $5 \times 10^3 / f = .1$ mm. This would mean that our (2) grid separation would have to be less than $.1/4 = .025$ mm in order to have propagation along the grid. However, for a practical case we might have an electrode mass of about 1% of the mass of the quartz under the electrode (1% "plateback"). The resonance frequency is somewhere between the infinite plate frequency of the unelectroded quartz and the infinite plate frequency of the electroded quartz. If we assume that it's half way - that is $f = 1.005f_0$ - then

$$v_z = \frac{5000}{\sqrt{1 - (1/1.005)^2}} = 50,000 \text{ M/Sec} \quad [A.9]$$

and the wavelength is 1 mm. If we choose a grid spacing of $< .25$ mm we have propagation on the grid, and for a grid spacing of $.125$ mm our effective velocity is 95% that of an arbitrarily small grid spacing (see Figure 2). This decrease in effective velocity alters the apparent electrode size by 5%, but electrode size is only a 2nd order dimensional effect in the calculation of resonance frequency. In other words, without presenting a formal proof, it appears as if a grid separation of the order of the plate thickness is adequate for reasonable results. It has been determined empirically that a minimum of 5 points in each lateral direction gives good results for the principal resonance and close-in spurious responses.

NONLINEAR ELECTROELASTIC EQUATIONS OF WAVE PROPAGATION AND VIBRATIONS IN QUARTZ BARS

M. Cengiz Dökmeci

Faculty of Aeronautics and Astronautics
Istanbul Technical University, P.K.9, Istanbul 80191

SUMMARY

This paper presents the nonlinear electroelastic equations of wave propagation and vibrations in a quartz bar of uniform cross-section. To begin with, Hamilton's principle is stated for a nonlinear elastic continuum with small piezoelectric coupling, and then by carrying out the pertinent variations, a variational principle with certain constraints is formulated. The constraints are incorporated into this principle through the dislocation potentials and Lagrange undetermined multipliers, and hence a generalized variational principle is derived for the motions of nonlinear piezoelectric continuum. Next, the generalized principle together with a series expansions of its mechanical displacements and electric potential, a system of nonlinear equations of the quartz bar is consistently obtained. These one-dimensional equations of higher orders of approximation in which account is taken of only the elastic nonlinearities govern all the types of extensional, flexural and torsional motions of quartz bar. Also, special motions of quartz bar and those of quartz bar with initial stresses are pointed out. Lastly, the fully linearized governing equations of quartz bar are considered, the uniqueness of their solutions is examined and the sufficient conditions are enumerated for the uniqueness.

1. INTRODUCTION

Essentially, piezoelectricity is a reversible, inherently anisotropic and polarizable but not magnetizable field, and the field is quasi-static and linear. In piezoelectricity, the elastic field is considered to be dynamic, while the electric field is taken to be static; this quasi-static approximation is well justified in all cases of engineering interest. Besides linearity in piezoelectricity, there may exist either an intrinsic nonlinearity or an induced nonlinearity. The former is peculiar to a piezoelectric material, whereas the latter is due to its deformation. The application of intrinsic or induced nonlinearity and/or both of them can significantly affect the mechanical behavior of piezoelectric elements. This is desirable in some cases and it has been examined only for a few particular motions.¹ In view of this review article,¹ the present paper is concerned in deriving the one-dimensional nonlinear electroelastic equations describing all the types of motion of thin cylindrical quartz bars.

Recently, extensive studies have been made of one-dimensional piezoelectric problems at low frequencies.^{2-5, 10-21} They have been directed toward either deriving differential governing equations of bars as in few cases^{2-5, 7-9} or solutions of specific bar problems as in most cases.¹⁰⁻²¹ Among the former cases, Milsom and his colleagues^{2,3} have presented a three-dimensional mode-matching theory of piezoelectric

rectangular quartz bar. Tiersten and Ballato⁴ have constructed the macroscopic equations accounting for the nonlinear extensional motion of thin piezoelectric rods, and they have applied these equations in the analyses of both intermodulation and nonlinear resonance of quartz rods. As a special case of their electromagnetic theory of rods, Green and Naghdi⁵ have studied the isothermal vibrations of piezoelectric crystal rods. Moreover, following Mindlin,⁶ the author⁷⁻⁹ has derived a one-dimensional theory of vibrations, which accommodates all the types of motions of piezoelectric crystal bars for both low as well as high frequencies. He has taken into account all the mechanical and electrical effects, and also he has described an application to biomechanics.

Efforts to solve certain problems of piezoelectric bars have been recently made by various authors.^{1,10-21}

Eer Nisse¹³ has calculated approximately the electrode stress effects for length-extensional and flexural resonant vibrations of long, thin bars of quartz. An analysis of the flexural-mode equation has been presented for a rod with a vibration isolator.¹⁴ The mechanical behavior of a piezoelectric bar has been studied with an electrical voltage as well as a time-dependent flux of heat at one end.¹⁵ A simple one-dimensional model has been used to investigate the effect of the relaxation time on the behavior of a semi-infinite piezoelectric rod under a thermal shock at its end.¹⁶ Moreover, the extensional vibration of a cylindrical rod with longitudinal piezoelectric coupling has been dealt with in an approximate procedure, and the depolarizing-field effect has been analyzed in a rod of finite and infinite length.¹⁷ A detailed numerical analysis of the dispersion relations has been reported for the axisymmetric normal waves of a piezoelectrically active bar waveguide.¹⁸ Further, the vibrational dissipation characteristics of a piezoceramic bar have been considered,¹⁹ as has the electrical excitation of an asymmetrically radiating bar.²⁰ Most recently, Solov'ev²¹ has examined the influence of the electroded zone on the natural frequency of thickness resonance of a piezoceramic rod of rectangular cross-section under the conditions of plain strain.

Our aim in the present paper is (i) to obtain a variational formulation for the nonlinear equations of an electroelastic solid with small piezoelectric coupling, with the help of this formulation, (ii) to derive a one-dimensional nonlinear electroelastic equations describing all the types of motions of thin quartz rods, and then (iii) to consider special motions of quartz bars and those of quartz bars with initial stresses, and also to examine the uniqueness of solutions in the linearized bar equations.

In the description of motions of the electroelastic solid, only the elastic nonlinearities are included, and hence the electrical behavior is taken to be linear. Accordingly, in the treatment of quartz rods which have small piezoelectric coupling, the nonlinear stress equations of motion, the associated nonlinear boundary conditions and the nonlinear strain-mechanical displacement relations are used, while the linear charge equations of electrostatics, the associated linear boundary conditions and alike are employed. Also, in the constitutive equations, the second-order, third-order and fourth-order elastic coefficients of quartz are retained for the stress tensor, and only the linear terms for the electric displacements.

Specifically, the content of this paper is as follows. First, the three-dimensional nonlinear equations of electroelastic solid are summarized in Section 2. This is followed in Section 3, by Hamilton's principle for the electroelastic solid and the associated quasi-variational principles. The geometry of a quartz bar is described, and also the series expansions for the mechanical displacements and the electric potential of quartz bar are recorded in Section 4. The nonlinear electroelastic equations of quartz bar are derived by means of the quasi-variational principles together with the series expansions in Section 5. Special motions of quartz bar are considered, and especially the linearized equations and the uniqueness in their solutions are studied in Section 6. Finally, the concluding remarks and further needs of research are indicated in Section 7.

Note - In this paper, standard Cartesian tensor notation is used in a Euclidean 3-space Ξ . The x_k -system of the space Ξ is identified with a fixed, right-handed system of Cartesian convected (intrinsic) coordinates. Einstein's summation convention is implied for all repeated Latin indices (1,2,3) and Greek indices (1,2), unless indices are enclosed with parantheses. Further, commas and primes stand for partial differentiations with respect to the indicated space coordinates and the coordinate x_3 , the bar axis, respectively, and superposed dots for time differentiations. Asterisks are used to designate prescribed quantities. The symbol $B(t)$ refers to a region B with its boundary surface ∂B and closure $\bar{B}(=B \cup \partial B)$, at time t in the space Ξ , and BXT refers to the Cartesian product of the region B and the time interval $T=[t_0, t_1]$. Also, boldface brackets are introduced so as to denote the jump of enclosed quantity across a surface of discontinuity S of the region B .

2. NONLINEAR PIEZOELECTRIC EQUATIONS

In the three-dimensional space Ξ , let $B \cup \partial B$ stand for an arbitrary, simply-connected, finite and bounded region of space occupied by an anisotropic elastic continuum with small piezoelectric coupling at time $t=t_0$. The regular boundary surface ∂B is consist of the complementary subsurfaces (S_u, S_t) and (S_σ, S_ϕ) , that is, $S_u \cup S_t = S_\sigma \cup S_\phi = \partial B$ and $S_u \cap S_t = S_\sigma \cap S_\phi = \emptyset$. Also, let $\bar{B}XT$ represent the domain of definitions for the functions of (x_k, t) .

Now, the three-dimensional differential equations are expressed for the electroelastic continuum with small piezoelectric coupling in the x_k -system of

Cartesian coordinates as follows.^{22,23}

Divergence Equations

$$t_{kl,k} - \rho a_l = 0 \quad \text{in } \bar{B}XT \quad (2.1)$$

$$t_{kl} = \tau_{kl} + T_{kl} = \tau_{kr}(\delta_{lr} + u_{l,r}) \quad (2.2)$$

$$D_{k,k} = 0 \quad \text{in } \bar{B}XT \quad (2.3)$$

with the definitions

t_{kl} = asymmetric Lagrangian stress tensor measured per unit area of the undeformed body

τ_{kl} = symmetric Kirchhoff stress tensor

$T_{kl} = \tau_{kr}u_{l,r}$ = Maxwell electrostatic stress tensor

ρ = density of the undeformed body

a_k = Lagrangian acceleration vector

u_k = displacement vector

δ_{kl} = Kronecker delta

D_k = electric displacement vector

Here, Eq.(2.1) stands for the nonlinear stress equations of motion and Eq.(2.3) for the linear charge equation of electrostatics. In Eqs. (2.1) and (2.3), when the stress tensor t_k per unit area of the undeformed body, associated with a surface in the deformed body, is referred to the base vectors in the deformed body, τ_{kl} arises, while if t_k is referred to the base vectors in the undeformed body, t_{kl} ensues.

Gradient Equations

$$S_{kl} = \frac{1}{2} (u_{k,l} + u_{l,k} + u_{r,k}u_{r,l}) \quad \text{in } \bar{B}XT \quad (2.4a)$$

$$S_{kl} = e_{kl} + \frac{1}{2} (e_{rk} + w_{rk})(e_{rl} + w_{rl}) \quad (2.4b)$$

$$e_{kl} = \frac{1}{2} (u_{k,l} + u_{l,k}), \quad w_{kl} = \frac{1}{2} (u_{k,l} - u_{l,k}) \quad (2.5)$$

$$E_k = -\phi_{,k} \quad \text{in } \bar{B}XT \quad (2.6)$$

with the definitions

S_{kl} = Lagrangian strain tensor

e_{kl} = linear strain tensor

w_{kl} = rotation tensor

ϕ_k = electric potential

E_k = quasi-static electric field vector

Equation (2.4) represents the nonlinear strain-mechanical displacement relations and Eq. (2.5) the electric field-electric potential relations.

Constitutive Equations

$$\tau_{kl} = \frac{1}{2} \left(\frac{\partial H}{\partial S_{kl}} + \frac{\partial H}{\partial S_{lk}} \right) \quad \text{in } \bar{B}XT \quad (2.7)$$

$$D_k = - \frac{\partial H}{\partial E_k} \quad \text{in } \bar{B}XT \quad (2.8)$$

with the definitions

$H = U - E_k D_k$ = electric enthalpy

U = potential energy density

A quartic form of the electric enthalpy is recorded in the form

$$H = \frac{1}{2} C_{klmn} S_{kl} S_{mn} - \frac{1}{2} C_{kl} E_k E_l - C_{klm} E_k S_{lm} + \frac{1}{6} C_{klmnr} S_{kl} S_{mn} S_{rs} + \frac{1}{24} C_{klmnrstu} S_{kl} S_{mn} S_{rs} S_{tu} \quad (2.9)$$

In view of Eqs. (2.7) and (2.8), this equation yields the nonlinear constitutive equations for the components τ_{kl} of the symmetric stress tensor and the linear constitutive equations for the components D_k of the electric displacement vector as

$$\tau_{kl} = C_{klmn} S_{mn} - C_{mkl} E_m + \frac{1}{2} C_{klmnr} S_{mn} S_{rt} + \frac{1}{6} C_{klmnrtpu} S_{mn} S_{rt} S_{pu} \quad (2.10)$$

$$D_k = C_{klm} S_{lm} + C_{kl} E_l \quad \text{in } B \times T \quad (2.11)$$

Here, C_{klmn} , C_{klmnr} and $C_{klmnrtpu}$ are the second-order, third-order and fourth-order elastic constants, C_{klm} is the piezoelectric strain constants and C_{kl} the dielectric permittivity. Of these constants, the elastic constants refer to free constants since they describe the stress-strain relations when the electric field is absent, while the remaining constants refer to clamped constants.²⁴ Further, the usual symmetry relations hold for these material constants, namely,

$$\begin{aligned} C_{klmn} &= C_{lkmn} = C_{mnkl}, \quad C_{klm} = C_{kml}, \quad C_{kl} = C_{lk} \\ C_{klmnr} &= C_{mnklr} = C_{rtmnkl} = C_{lkmnr} \\ C_{klmnrtpu} &= C_{mnklrtpu} = C_{rtmnklpu} = C_{pumnrtkl} \\ &= C_{lkmnrtpu} \end{aligned} \quad (2.12)$$

Boundary Conditions

$$t_k^* - n_l t_{lk} = t_k^* - n_l \tau_{lr} (\delta_{kr} + u_{k,r}) = 0 \quad \text{on } S_t \times T \quad (2.13)$$

$$\sigma^* - n_k D_k = 0 \quad \text{on } S_\sigma \times T \quad (2.14)$$

$$u_k - u_k^* = 0 \quad \text{on } S_u \times T \quad (2.15)$$

$$\phi - \phi^* = 0 \quad \text{on } S_\phi \times T \quad (2.16)$$

with the definitions,

$$\begin{aligned} t_k &= n_l t_{lk} = \text{stress vector} \\ n_k &= \text{outward unit vector normal to } \partial B \\ \sigma &= n_k D_k = \text{surface charge} \end{aligned}$$

Initial Conditions

$$\begin{aligned} u_k(x_1, t_0) - u_k^*(x_1) &= 0 \\ \dot{u}_k(x_1, t_0) - \dot{u}_k^*(x_1) &= 0 \quad \text{in } B(t_0) \\ \phi(x_k, t_0) - \phi^*(x_k) &= 0 \end{aligned} \quad (2.17)$$

Jump Conditions

$$n_k \Pi t_{kl} \Pi = 0 \quad (2.18a)$$

$$n_k \Pi \tau_{kr} (\delta_{lr} + u_{l,r}) \Pi + t_1^a = 0 \quad (2.18b)$$

$$n_k \Pi D_k \Pi = Q \quad \text{on } S \times T \quad (2.19)$$

$$\Pi u_k \Pi = 0 \quad (2.20)$$

$$\Pi \phi \Pi = 0 \quad (2.21)$$

with the definitions

$$\begin{aligned} t_1^a &= \text{applied prescribed surface traction} \\ Q &= \text{electric surface charge density} \\ S &= \text{material surface of discontinuity} \end{aligned}$$

Equations (2.1)-(2.17) completely describe the nonlinear behavior of electroelastic continuum with small piezoelectric coupling, and the last four equations arise at a material surface of discontinuity.

3. VARIATIONAL FORMULATION

In piezoelectricity, the fundamental equations have been often expressed in variational forms as the appropriate Euler equations of variational principles.^{1,25-30} These variational principles have been primarily derived with the aid of Hamilton's

principle^{1,6,25}, and they allow the establishment of lower order theories and approximate direct solutions in piezoelectricity.^{26,31} Now, Hamilton's principle is stated for the nonlinear elastic continuum with small piezoelectric coupling as

$$\delta H = \delta \int \mathcal{L} dt + \int \delta W dt = 0 \quad (3.1)$$

with the definitions

$$\mathcal{L} = \int_B [K - H(S_{kl}, E_k)] dV \quad (3.2)$$

$$K = \frac{1}{2} \rho \dot{u}_k \dot{u}_k$$

$$\delta W = \int_{\partial B} (t_k^* \delta u_k - \sigma^* \delta \phi) dS \quad (3.4)$$

where \mathcal{L} is the lagrangian function, K the kinetic energy density and δW the virtual work per unit area done by the prescribed surface tractions t_k^* and surface charge σ^* .

By inserting Eqs. (3.2)-(3.4) into Eq. (3.1), one arrives at the variational equation of the form.

$$\begin{aligned} \delta H &= \delta \int dt \left\{ \int_B \left[\frac{1}{2} \rho \dot{u}_k \dot{u}_k - H(S_{kl}, E_k) \right] dV \right. \\ &\quad \left. + \int_{\partial B} (t_k^* \delta u_k - \sigma^* \delta \phi) dS \right\} = 0 \end{aligned} \quad (3.5)$$

where all variations vanish at $t=t_0$ and $t=t_1$. Taking the indicated variations, utilizing the fact that the operation of variation commutes with that of differentiation, integrating by parts with respect to time and employing the constitutive relations (2.7) and (2.8) and the constraints on the variations, Eq. (3.5) takes the form

$$\begin{aligned} \delta H &= - \int_B dt \int (\rho a_k \delta u_k + \tau_{kl} \delta S_{kl} - D_k \delta E_k) dV \\ &\quad + \int_{\partial B} dt \int (t_k^* \delta u_k - \sigma^* \delta \phi) dS = 0 \end{aligned}$$

By substituting the nonlinear strain-mechanical displacement relations (2.4) and the linear electric field-electric potential relations into this equation, employing the divergence theorem and rearranging terms, one finally obtains

$$\begin{aligned} \delta H &= \int_B dt \int \{ [\tau_{kr} (\delta_{lr} + u_{l,r})]_{,k} - \rho a_1 \} \delta u_1 dV \\ &\quad + \int_B dt \int D_{k,k} \delta \phi dV \\ &\quad + \int_{\partial B} dt \int [t_k^* - n_l \tau_{lr} (\delta_{kr} + u_{k,r})] \delta u_k dS \\ &\quad + \int_{\partial B} dt \int (\sigma^* - n_k D_k) \delta \phi dS = 0 \end{aligned} \quad (3.6)$$

In deriving this variational principle, the principle of conservation of mass is considered and the condition $\delta u_k = \delta \phi = 0$ in $B(t_0)$ and $B(t_1)$ (3.7)

is imposed. Since the variations δu_k and $\delta \phi$ of the admissible state $\Lambda_h = \{u_k, \phi\}$ in Eq. (3.6) are arbitrary and independent inside the volume B and on the boundary surface ∂B , one has the nonlinear stress equations of motion (2.1), the linear charge equation of electrostatics (2.3) and the associated natural boundary conditions of tractions and surface charge (2.13) and (2.14) as the appropriate Euler equations of the variational principle (3.6). This is a two-field variational principle, and it contains some of earlier variational principles as special cases.^{1,25,32,33} Further, it is of interest to note that this variational principle can be extracted from the principle of virtual work as well.³⁰

The differential variational principle (3.6) can

be used, as usual, in solving approximately the boundary-value problems of nonlinear elastic continuum with small piezoelectric coupling, provided that the initial conditions (2.17) may be left out of account by

a variety of numerical techniques.^{34,35} Besides, any approximating solution must satisfy the rest of the fundamental equations (2.4) and (2.6)-(2.8) in accordance with Eq. (3.6) as well as the usual continuity and differentiability conditions of field quantities and the condition that the stress tensor be symmetric. This feature of Hamilton's principle has been discussed very thoroughly by Tiersten³⁶ and Gurtin.³⁷ However, the constraint conditions (2.4) and (2.6)-(2.8) can be relaxed through certain methods.³⁷⁻³⁹ Of these methods, Friedrichs's transformation³⁸ is used herein so as to remove the constraint conditions due to

its versatility and wide use in the literature.³⁹ Accordingly, to adjoin the constraint conditions into the quasi-variational principle (3.6), the dislocation potentials $\Delta\alpha\beta$, each constraints as a zero times a Lagrange multiplier, are introduced as

$$\begin{aligned}\Delta_{11} &= \lambda_{k1} [S_{k1} - \frac{1}{2}(u_{k,1} + u_{1,k} + u_{r,k}u_{r,1})] \\ \Delta_{12} &= \lambda_k (u_k - u_k^*) \\ \Delta_{21} &= \mu(\phi - \phi^*) \\ \Delta_{22} &= \mu_k (E_k + \phi, k)\end{aligned}\quad (3.8)$$

and they are added to Eq. (3.1), namely,

$$\delta J = \delta H + \delta \int_B \{ \int_B (\Delta_{11} + \Delta_{22}) dV + \int_{S_u} \Delta_{12} dS + \int_{S_\phi} \Delta_{21} dS \} = 0 \quad (3.9)$$

with the virtual work of the form

$$\delta W = \int_{S_t} t_k^* \delta u_k dS - \int_{S_\sigma} \sigma^* \delta \phi dS \quad (3.10)$$

Then, treating all the variations in Eq. (3.9) as free, one finds

$$\begin{aligned}\delta J &= \int_B \{ dt \int_B [- \rho a_k \delta u_k - \frac{1}{2} (\frac{\partial H}{\partial S_{k1}} + \frac{\partial H}{\partial S_{1k}}) \delta S_{k1} - \frac{\partial H}{\partial E_k} \delta E_k] dV \\ &+ \int_B \{ dt \int_{S_t} t_k^* \delta u_k - \int_{S_\sigma} \sigma^* \delta \phi dS \\ &+ \int_B \{ dt \int_B \{ \delta \lambda_{k1} [S_{k1} - \frac{1}{2}(u_{k,1} + u_{1,k} + u_{r,k}u_{r,1})] \\ &+ \lambda_{k1} [\delta S_{k1} - (\delta u_{k,1} + u_{r,k} \delta u_{r,1})] \} dV \\ &+ \int_B \{ dt \int_B [\delta \mu_k (E_k + \phi, k) + \mu_k (\delta E_k + \delta \phi, k)] dV \\ &+ \int_{S_u} \{ dt \int_{S_u} [\delta \lambda_k (u_k - u_k^*) + \lambda_k \delta u_k] dS \\ &+ \int_{S_\phi} \{ dt \int_{S_\phi} [\delta \mu (\phi - \phi^*) + \mu \delta \phi] dS = 0 \end{aligned}\quad (3.11)$$

As before, by applying the divergence theorem and after some rearrangement, the stationary condition (3.11) readily gives the Lagrange multiplier in the form

$$\begin{aligned}\lambda_{k1} &= \tau_{k1} \quad , \quad \mu_k = - D_k \quad \lambda_k = t_k = n_1 t_{1k} \\ \mu &= - \sigma = - n_k D_k\end{aligned}\quad (3.12)$$

since the volumetric variations δu_k , $\delta \phi$, δS_{k1} , δE_k , $\delta \lambda_{k1}$ and $\delta \mu_k$ are arbitrary and independent in the region B and the surface variations δu_k , $\delta \phi$, $\delta \lambda_k$ and $\delta \mu$ on the boundary surfaces S_u , S_ϕ , S_t and S_σ .

Finally, from Eqs. (3.9), (3.10) and (3.12), one obtains the variational principle as follows.

$$\delta J \langle \Lambda \rangle = \int_B \delta J_{k1} dV = 0 \quad (3.13a)$$

where

$$\Lambda = \{ u_k, t_k, \tau_{k1}, S_{k1}; \phi, \sigma, D_k, E_k \} \quad (3.13b)$$

and

$$\delta J_{1111} = \int_B \{ [\tau_{kr} (\delta_{1r} + u_{1,r})]_{,k} - \rho a_1 \} \delta u_1 dV \quad (3.14)$$

$$\delta J_{1212} = \int_B D_{k,k} \delta \phi dV \quad (3.15)$$

$$\delta J_{1313} = \int_B [\tau_{k1} - \frac{1}{2} (\frac{\partial H}{\partial S_{k1}} + \frac{\partial H}{\partial S_{1k}})] \delta S_{k1} dV \quad (3.16)$$

$$\delta J_{2121} = - \int_B (D_k + \frac{\partial H}{\partial E_k}) \delta E_k dV \quad (3.17)$$

$$\delta J_{2222} = \int_B [S_{k1} - \frac{1}{2} (u_{k,1} + u_{1,k} + u_{r,k}u_{r,1})] \delta \tau_{k1} dV \quad (3.18)$$

$$\delta J_{2323} = - \int_B (E_k + \phi, k) \delta D_k dV \quad (3.19)$$

$$\delta J_{3131} = \int_{S_t} [t_k^* - n_1 \tau_{1r} (\delta_{kr} + u_{k,r})] \delta u_k dS \quad (3.20)$$

$$\delta J_{3232} = \int_{S_\sigma} (\sigma^* - n_k D_k) \delta \phi dS \quad (3.21)$$

$$\delta J_{3333} = - \int_{S_u} (u_k^* - u_k) \delta t_k dS - \int_{S_\phi} (\phi^* - \phi) \delta \sigma dS \quad (3.22)$$

This variational principle may be written in a compact form by

$$\begin{aligned}\delta J \langle \Lambda \rangle &= \delta \int_B \{ \int_B \{ \tau_{k1} [S_{k1} - \frac{1}{2} (u_{k,1} + u_{1,k} + u_{r,k}u_{r,1})] \\ &- D_k (E_k + \phi, k) + K - H (S_{k1}, E_k) \} dV \\ &- \int_{S_u} (u_k^* - u_k) t_k dS + \int_{S_t} t_k^* u_k dS \\ &- \int_{S_\phi} (\phi^* - \phi) \sigma dS + \int_{S_\sigma} \sigma^* \phi dS \} = 0 \end{aligned}\quad (3.23)$$

The variational equation (3.13) or (3.23) generates, as its Euler equations, the fundamental equations of nonlinear elastic continuum with small piezoelectric coupling, and hence we conclude a variational principle below.

V a r i a t i o n a l P r i n c i p l e : Let $B \cup \partial B$ denote a regular, finite and bounded region of the space \mathbb{R}^3 , with its piecewise smooth boundary surface $\partial B (= S_u \cup S_t \cup S_\phi \cup S_\sigma)$ and $S_u n_t = S_\phi n_\sigma = 0$ and its closure

$\bar{B} (= B \cup \partial B)$. Then, of all the admissible states $\Lambda (= u_k, t_k, \tau_{k1}, S_{k1}; \phi, \sigma, D_k, E_k)$ which satisfy the initial conditions (2.17) as well as the symmetry of stress tensor τ_{k1} and the usual continuity and

differentiability conditions of field variables; if and only if, that admissible state Λ which satisfies the nonlinear stress equations of motion (2.1), the linear charge equation of electrostatics (2.3), the nonlinear strain-mechanical displacement relations (2.4), the electric field-electric potential relations (2.6), the nonlinear constitutive equations (2.7) and (2.8), and the natural boundary conditions (2.13)-(2.16), is determined by the variational equation $\delta J \langle \Lambda \rangle = 0$ of Eq. (3.23) as its appropriate Euler equations.

The variational principle (3.23) is believed to be first reported herein, and it does agree with and

represents, as special cases, certain earlier variational principles operating only on some of the field variables.^{25,30,32,33} By use of the fundamental lemma of the calculus of variations, the principle (3.13) or (3.23) leads readily to all the fundamental equations of piezoelectric continuum with small piezoelectric coupling, Eqs. (2.1), (2.3), (2.4), (2.6)-(2.8) and (2.13)-(2.16), but the initial conditions (2.17); conversely, if these equations are met, the principle is obviously satisfied. Further, the variational principle can be readily expressed, following Tiersten,³⁶ in an augmented form which incorporates the initial conditions as well as the jump conditions (2.18)-(2.21); the result is a differential variational principle.^{29,30}

In closing, it is of interest to note that the expressions δJ_{1313} and δJ_{2121} in Eqs. (3.16) and (3.17) take the form

$$\begin{aligned} \delta \bar{J}_{1313} = & \int_B [\tau_{kl} - (C_{klmn} S_{mn} - C_{mkl} E_m \\ & + \frac{1}{2} C_{klmnr} S_{mn} S_{rt} \\ & + \frac{1}{6} C_{klmnrtpu} S_{mn} S_{rt} S_{pu})] \delta S_{kl} dV \quad (3.26) \end{aligned}$$

$$\delta \bar{J}_{2121} = - \int_B [D_k - (C_{klm} S_{lm} + C_{kl} E_l)] \delta E_k dV \quad (3.27)$$

in the case when the constitutive equations (2.10) and (2.11) are used in lieu of Eqs. (2.7) and (2.8).

4. EXPANSION IN SERIES

This section deals with the description of bar geometry, the method of reduction in deriving the electroelastic equations of quartz bar and the expansion in series for the field variables of quartz bar.

Geometry of Quartz Bar - Consider an initially slender quartz bar of uniform cross-section in the Euclidean 3-space Ξ . The bar is referred to a system of right-handed Cartesian convected coordinates x_k . The axes x_α are selected as the principal axes of bar cross-section, and the axis x_3 is taken as the locus of centroids of bar cross-sections which is a straight line in the undeformed bar. The cross-section of bar A is bounded by a simply-connected Jordan curve C, that is, sufficiently smooth and non-intersecting. Moreover, by definition, one has the fundamental assumption of bars, $d/l \ll 1$, where d is the maximum diameter of cross-section and l is the length of quartz bar. In addition to this, no singularities of any type is supposed to be present within the region of quartz bar. Thus, the bar is treated as a one-dimensional continuous model of a three-dimensional body.

Method of Reduction - The presence of electric field and material anisotropy makes it almost always compulsory the use of approximate lower order equations in investigating the dynamic characteristics of piezoelectric elements. Of the standard techniques^{6, 40-43} to reduce the three-dimensional equations of piezoelectricity into the lower order equations,

Mindlin's method of reduction⁶ is especially suitable and wide use in the literature,^{25,40-43} and it is used herein so as to construct the nonlinear electroelastic equations of quartz bar. This method of reduction rests entirely on the series expansions of field variables which are inserted in a pertinent variational principle which is then integrated with respect to one-

or two-dimension.

Expansion in Power Series - Under the usual existence, regularity and smoothness assumptions of bars and their fundamental assumptions, mentioned above, a set of shape functions $(\beta_{11}, \beta_{12}, \dots, \beta_{mn})$ is selected, and the shape functions are taken to be complete for a given field quantity in the bar region. Then the electric potential and the displacement components are represented by

$$\{\phi, u_k\} = \sum_{m,n=0}^{N=\infty} \{\phi^{(m,n)}(x_3, t), u_k^{(m,n)}(x_3, t)\} \beta_{mn}(x_1, x_2) \quad (4.1a)$$

Here, $\phi^{(m,n)}$ and $u_k^{(m,n)}$ are unknown *a priori* and independent functions of electric potential and mechanical displacements of order (m,n) to be determined, and the shape functions β_{mn} of order (m,n) can be selected to be any type of functions which is appropriate to the contour of cross-section and they are taken as a power series of the form

$$\beta_{mn}(x_\alpha) = x_1^m x_2^n \quad (4.1b)$$

in the present analysis.

5. NONLINEAR BAR EQUATIONS

In this section, by means of the method of reduction described in Section 4, the system of one-dimensional, nonlinear electroelastic equations of quartz bar is consistently derived. To begin with, the series expansions (4.1) are inserted into the variational principle (3.13), the volume integrals are split into an area integral over a cross-section of, and a line integral along, the quartz bar, and then the integrations are performed. The resulting equations are recorded below in terms of various field quantities of higher orders which are now defined.

Field Quantities of Order (m,n) - The stress resultants of order (m,n) :

$$T_{kl}^{(m,n)} = \int_A x_1^m x_2^n \tau_{kl} dA \quad (5.1)$$

$$\begin{aligned} N_k^{(m,n)} = & \sum_{p+q=0}^N \{ [mpT_{11}^{(m+p-2, n+q)} \\ & + (np+mq)T_{12}^{(m+p-1, n+q-1)} + qnT_{22}^{(m+p, n+q-1)} \\ & + pT_{31}^{(m+p-1, n+q)} + qT_{32}^{(m+p, n+q-1)}] u_k(p, q) \\ & + T_{33}^{(m+p, n+q)} u_k'(p, q) + [(p+m)T_{13}^{(m+p-1, n+q)} \\ & + (q+n)T_{23}^{(m+p, n+q-1)} + T_{33}^{(m+p, n+q-1)}] u_k'(p, q) \} \end{aligned} \quad (5.2)$$

the surface loads of order (m,n) :

$$P_k^{(m,n)} = \oint_C x_1^m x_2^n \gamma_\alpha \tau_{\alpha k} dA \quad (5.3)$$

$$Q_k^{(m,n)} = P_k^{(m,n)} + R_k^{(m,n)} \quad (5.4)$$

$$\begin{aligned} R_k^{(m,n)} = & \sum_{p+q=0}^N [(pP_1^{(m+p-1, n+q)} \\ & + qP_2^{(m+p, n+q-1)}) u_k(p, q) + P_3^{(m+p, n+q)} u_k'(p, q)] \quad (5.5) \end{aligned}$$

$$\begin{aligned} N_{3k}^{(m,n)} = & \sum_{p+q=0}^N [(pT_{31}^{(m+p-1, n+q)} \\ & + qT_{32}^{(m+p, n+q-1)}) u_k(p, q) \\ & + T_{33}^{(m+p, n+q)} u_k'(p, q)] \quad (5.6) \end{aligned}$$

the acceleration of order (m,n) :

$$U_k(m,n) = \sum_{p+q=0}^N I^{(m+p, n+q)} u_k(p,q) \quad (5.7a)$$

$$A_k(m,n) = U_k(m,n) \quad (5.7b)$$

the prescribed stress resultants of order (m,n):

$$T_k^*(m,n) = \int_A x_1^m x_2^n t_k^* dA, \quad P_k^*(m,n) = \oint_C x_1^m x_2^n t_k^* ds \quad (5.8)$$

the aerial moment of inertia of order (m,n):

$$I^{(m,n)} = \int_A x_1^m x_2^n dA \quad (5.9a)$$

with the usual quantities of bars as

$$I^{(0,0)} = A, \quad I^{(1,0)} = I^{(0,1)} = I^{(1,1)} = 0 \quad (5.9b)$$

the electric displacements of order (m,n):

$$D_k^{(m,n)} = \int_A x_1^m x_2^n D_k dA \quad (5.10)$$

the surface charge of order (m,n):

$$D^{(m,n)} = \oint_C x_1^m x_2^n v_\alpha D_\alpha ds \quad (5.11)$$

and the prescribed surface charge of order (m,n):

$$\phi^*(m,n) = \int_A x_1^m x_2^n \sigma^* dA, \quad D^*(m,n) = \oint_C x_1^m x_2^n \sigma^* ds \quad (5.12)$$

are defined. In the above equations, v_α denotes the unit outward vectors normal to the contour C of cross-section. Also, the electric enthalpy function G measured per unit length of the undeformed bar, namely,

$$G = \int_A H dA \quad (5.13)$$

is introduced for later convenience.

Equations of Motion - Consider the volume integral (3.14) of the form, namely,

$$\delta J_{1111} = \int_0^1 dx_3 \{ [\tau_{kr}(\delta_{1r} + u_{1,r})]_{,k} - \rho a_1 \} \delta u_1 dV \quad (5.14)$$

Substituting the series expansions of mechanical displacements (4.1) into this integral, carrying out the integrations over A, using the divergence theorem and replacing the stress and load resultants of order (m,n), one obtains

$$\begin{aligned} \delta J_{1111} = & \int_L dx_3 \sum_{m+n=0}^N (T_{3k}^{(m,n)} - mT_{1k}^{(m-1,n)} \\ & - nT_{2k}^{(m,n-1)} + N_k^{(m,n)} + Q_k^{(m,n)} \\ & - \rho A_k^{(m,n)}) \delta u_k^{(m,n)} \end{aligned} \quad (5.15)$$

where L stands for the interval [0,1]. When setting the variational equation (3.13) equal to zero for the arbitrary and independent variations of field quantities such as $\delta u_k^{(m,n)}$ in this case, one readily obtains the macroscopic equations of motion of order (m,n) in the form

$$\begin{aligned} T_{3k}^{(m,n)} - mT_{1k}^{(m-1,n)} - nT_{2k}^{(m,n-1)} + N_k^{(m,n)} \\ + Q_k^{(m,n)} - \rho A_k^{(m,n)} = 0 \quad \text{on } L \times T \end{aligned} \quad (5.16)$$

for the quartz bar.

Charge Equation of

Electrostatics - As before, evaluating the volume integral δJ_{1212} of Eq. (3.15), one arrives at the macroscopic charge equation of electrostatics of order (m,n) as

$$D_3^{(m,n)} - mD_1^{(m-1,n)} - nD_2^{(m,n-1)} + D^{(m,n)} = 0 \quad (5.17)$$

in terms of the quantities defined by Eqs. (5.10) and (5.11).

Electric Field and Strain Distributions - Likewise, considering Eqs. (3.18) and (3.19), integrating over A and using the stress and electric displacements of order (m,n), the distribution of strain of order (m,n):

$$S_{kl}(x_m, t) = \sum_{m+n=0}^N x_1^m x_2^n S_{kl}^{(m,n)}(x_3, t) \quad (5.18)$$

where

$$\begin{aligned} S_{kl}^{(m,n)} = & e_{kl}^{(m,n)} + \frac{1}{2} \sum_{p+q=0}^{m+n} (e_{rk}^{(m-p, n-q)} \\ & + w_{rk}^{(m-p, n-q)})(e_{rl}^{(p,q)} + w_{rl}^{(p,q)}) \end{aligned} \quad (5.19)$$

with

$$\begin{aligned} e_{\alpha\beta}^{(m,n)} = & \frac{1}{2} [(m+1)(\delta_{1\alpha} u_\beta^{(m+1,n)} + \delta_{1\beta} u_\alpha^{(m+1,n)} \\ & + (n+1)(\delta_{2\alpha} u_\beta^{(m,n+1)} + \delta_{2\beta} u_\alpha^{(m,n+1)})] \\ e_{\alpha 3}^{(m,n)} = & \frac{1}{2} [u_\alpha^{(m,n)} + (m+1)\delta_{1\alpha} u_3^{(m+1,n)} \\ & + (n+1)\delta_{2\alpha} u_3^{(m,n+1)}] \\ w_{\alpha\beta}^{(m,n)} = & \frac{1}{2} [(m+1)(\delta_{1\beta} u_\alpha^{(m+1,n)} - \delta_{1\alpha} u_\beta^{(m+1,n)} \\ & + (n+1)(\delta_{2\beta} u_\alpha^{(m,n+1)} - \delta_{2\alpha} u_\beta^{(m,n+1)})] \\ w_{\alpha 3}^{(m,n)} = & \frac{1}{2} [u_\alpha^{(m,n)} - (m+1)\delta_{1\alpha} u_3^{(m+1,n)} \\ & - (n+1)\delta_{2\alpha} u_3^{(m,n+1)}] \\ e_{33}^{(m,n)} = & u_3^{(m,n)}, \quad w_{33}^{(m,n)} = 0 \end{aligned} \quad (5.20)$$

and that of electric field of order (m,n):

$$E_k(x_1, t) = \sum_{m+n=0}^N x_1^m x_2^n E_k^{(m,n)}(x_3, t) \quad (5.21)$$

where

$$\begin{aligned} E_\alpha^{(m,n)} = & - [(m+1)\delta_{1\alpha} \phi^{(m+1,n)} \\ & + (n+1)\delta_{2\alpha} \phi^{(m,n+1)}] \\ E_3^{(m,n)} = & - \phi^{(m,n)} \end{aligned} \quad (5.22)$$

are found for the vanishing of the coefficients of free variations of the stress resultants and electric displacements of order (m,n) of quartz bar in the variational equation (3.13).

Constitutive Equations - Paralleling to the derivation of electric field and strain distributions above, the volume integrals (3.16) and (3.17) are evaluated by use of Eqs. (5.1), (5.10), (5.18) and (5.21), and then the constitutive relations are obtained for the stress resultants of order (m,n) and the electric displacements of order (m,n) in the form.

$$T_{kl}^{(m,n)} = \frac{1}{2} \left(\frac{\partial G}{\partial S_{kl}^{(m,n)}} + \frac{\partial G}{\partial S_{lk}^{(m,n)}} \right) \text{ on } L \times T \quad (5.23)$$

$$D_k^{(m,n)} = - \frac{\partial G}{\partial E_k^{(m,n)}} \quad \text{on } L \times T \quad (5.24)$$

in terms of the electric enthalpy function G of Eq. (5.13).

In the case of the linear constitutive equations (2.10) and (2.11), the volume integrals (3.26) and (3.27) are evaluated in lieu of Eqs. (3.16) and (3.17) with the result.

$$T_{kl}^{(m,n)} = c_{klrt} \sum_{a+b=0}^N I^{(m+a, n+b)} S_{rt}^{(a,b)}$$

$$\begin{aligned}
& - C_{rk} l \sum_{a+b=0}^N I(m+a, n+b) E_r^{(a,b)} \\
& + \frac{1}{2} C_{klrtpq} \sum_{a+b=0}^N \sum_{c+d=0}^N \chi_{rt}^{(a,b)} S_{pq}^{(c,d)} \quad (5.25) \\
& + \frac{1}{6} C_{klrtpquv} \sum_{a+b=0}^N \sum_{c+d=0}^N \sum_{e+f=0}^N \zeta_{rt}^{(a,b)} S_{pq}^{(c,d)} S_{uv}^{(e,f)}
\end{aligned}$$

with

$$\chi = I(m+a+c, n+b+d), \quad \zeta = I(m+a+c+e, n+b+d+f)$$

and

$$\begin{aligned}
D_k^{(m,n)} &= C_{klr} \sum_{a+b=0}^N I(m+a, n+b) S_{lr}^{(a,b)} \\
&+ C_{kl} \sum_{a+b=0}^N I(m+a, n+b) E_l^{(a,b)} \quad (5.26)
\end{aligned}$$

in terms of $I(m,n)$ of Eq. (5.8).

Boundary Conditions - The mechanical displacements and the surface charge are prescribed on the surface portion S_d of the lateral surface of bar S_l and on the left face boundary surface A_l , while the traction and the electric potential are prescribed on the remaining portion S_r of S_l and on the right face boundary surface A_r , where $S_u = S_d \cup A_l$, $S_t = S_r \cup A_r$, $S_d \cup S_r = S_l$, $A_l \cup A_r = \partial B$. As in the derivation of the stress equations of motion and the charge equation of electrostatics, by evaluating the surface integrals (3.20)-(3.22) of the variational principle (3.13), the natural boundary conditions are expressed for the tractions of order (m,n) by

$$p_k^{*(m,n)} - (p_k^{(m,n)} + R_k^{(m,n)}) = 0 \quad \text{on } S_d \times T \quad (5.27)$$

$$T_k^{*(m,n)} - (T_k^{(m,n)} + N_{3k}^{(m,n)}) = 0 \quad \text{on } A_l \times T \quad (5.28)$$

for the surface charge of order (m,n) by

$$D^{*(m,n)} - D^{(m,n)} = 0 \quad \text{on } S_d \times T \quad (5.29)$$

$$\phi^{*(m,n)} + D_3^{(m,n)} = 0 \quad \text{on } A_l \times T \quad (5.30)$$

for the mechanical displacements of order (m,n) by

$$u_k^{*(m,n)} - u_k^{(m,n)} = 0 \quad \text{on } S_u \times T \quad (5.31)$$

and for the electric potential of order (m,n) by

$$\phi^{*(m,n)} - \phi^{(m,n)} = 0 \quad \text{on } S_\phi \times T \quad (5.32)$$

Here, $(t_k$ and $\phi)$ and $(u_k$ and $\sigma)$ are prescribed, since they are the most commonly encountered in practice.⁷

Initial Conditions - By making use of Eq. (2.17) and Eq. (4.1), one reads the initial conditions of order (m,n) as

$$u_k^{(m,n)}(x_3, t) - v_k^{*(m,n)}(x_3) = 0 \quad \text{on } L(t_0) \quad (5.33)$$

$$\dot{u}_k^{(m,n)}(x_3, t) - w_k^{*(m,n)}(x_3) = 0 \quad \text{on } L(t_0) \quad (5.34)$$

and

$$\phi^{(m,n)}(x_3, t) - \psi^{*(m,n)}(x_3) = 0 \quad \text{on } L(t_0) \quad (5.35)$$

where v_k^* , w_k^* and ψ^* are given functions of x_3 .

Thus far, the one-dimensional, nonlinear equations of successively higher orders of approximation are consistently developed for quartz bars on the basis of three-dimensional theory of piezoelectricity. These governing equations of order (m,n) consist of the electric potential and mechanical displacement fields (4.1), the stress equations of motion (5.16), the charge equation of electrostatics (5.17), the electric field and strain distributions (5.19) and (5.21), the constitutive equations (5.23) and (5.24) or (5.25) and

(5.26), the natural boundary conditions (5.27)-(5.32) and the initial conditions (5.33)-(5.35). The number of the governing equations is infinite, that is, $m+n=0,1,2,\dots, N=\infty$, and hence the equations are not formally determinate yet; they will be made deterministic in the next section.

6. SPECIAL MOTIONS

To obtain a deterministic system of nonlinear electroelastic equations of quartz bar derived in the previous section, these infinite number of equations with their infinite number of unknowns must be consistently reduced to a finite number of equations with their finite number of unknowns by a process of series truncation. The process of truncation, special motions of quartz bar, and especially the linearized governing equations and the uniqueness in their solutions are taken up in this section. Further, the motions of quartz bar with initial stresses are pointed out.

Deterministic Bar Equations

The foregoing derivation of the governing equations of quartz bar rests entirely on the fields of mechanical displacements and electric potential, chosen *a priori* as a starting point and representing them by the power series expansions (4.1) of which the terms $u_k^{(m,n)}$ and $\phi^{(m,n)}$ are already taken to be exist. Thus, the governing equations of order (M,N) is defined by either

$$\{\phi, u_k\} = \sum_{m=0}^M \sum_{n=0}^N x_1^m x_2^n \{\phi^{(m,n)}, u_k^{(m,n)}\} \quad (6.1a)$$

or the series expansions (4.1) together with the condition

$$\phi^{(m,n)} = u_k^{(m,n)} = 0 \quad \text{for all } m \geq M+1, n \geq N+1 \quad (6.1b)$$

and only the quantities involved in (6.1) are kept in the equations. In view of Eqs. (6.1), there exists the $4(M+1)(N+1)$ unknowns $u_k^{(m,n)}$ and $\phi^{(m,n)}$ and equations to solve them. In addition to Eqs. (6.1), another type of deterministic governing equations is simply defined by

$$\phi^{(m,n)} = u_k^{(m,n)} = 0 \quad \text{for all } (m+n) \geq N+1 \quad (6.2)$$

where N is a positive integer. This obviously considers the same weight for both of the lateral coordinates x_1 and x_2 . Further, in both the definitions (6.1) and (6.2), by selecting the positive integers M and N or only N for particular applications, the governing equations incorporate as many higher order effects as deemed necessary. Hence the customary correction factors of bars are naturally abrogated.⁶

Special Motions

Of the special motions of quartz bar, the extensional motions^{4,4} can be examined by representing the electric potential and the mechanical displacements as in Eq. (4.1) with the condition $u_\alpha = u_\alpha(x_1, x_2, t) = 0$. Also, in the case of

low-frequency extensional motions, it is appropriate to take the vanishing boundary stresses on the lateral boundary surface S_l , and hence all the vanishing stresses but T_{33} . The electrical boundary conditions depend on the surface S_ϕ , and if the edge boundary surfaces $S_e (=A_l \cup A_r)$ are fully electroded, the boundary conditions become $D_{\alpha 3} = 0$ in Eqs. (5.29) and (5.11); this will be reported later.^{4,5} Moreover, the governing equations of quartz bar can be specialized to study its nonlinear torsional motions in the sense of Saint-Venant by the use of the displacement field (4.1)

together with the condition⁴⁶

$$u_1 = x_2 u_1^{(0,1)}, \quad u_2 = x_1 u_2^{(1,0)}, \quad u_3^{(m,n)} = \omega C_{mn} \quad (6.3a)$$

and

$$u_1^{(0,1)} = -u_2^{(1,0)} = -\omega x_3 \quad (6.3b)$$

Here, $\omega = \omega_{12}$ denotes the uniform rate of twist and C_{mn} is a constant.

Linear Bar Equations - Dropping out all the terms involving nonlinearity, namely,

$$W_{kl}^{(m,n)} = 0, \quad N_k^{(m,n)} = R_k^{(m,n)} = N_{3k}^{(m,n)} = 0 \\ I_{(m+a+c, n+b+d)} = I_{(m+a+c+e, n+b+d+f)} = 0 \quad (6.4)$$

in the macroscopic electroelastic equations of Section 5, the fully linear governing equations of quartz bar are obtained. They are the macroscopic equations of motion as

$$\begin{aligned} T_{3k}^{(m,n)} - mT_{1k}^{(m,n)} - nT_{2k}^{(m,n-1)} + P_k^{(m,n)} \\ - \rho A_k^{(m,n)} = 0 \quad \text{on } L \times T \end{aligned} \quad (6.5)$$

the associated boundary conditions of tractions as

$$P_k^{(m,n)} - p_k^{(m,n)} = 0 \quad \text{on } S_d \times T \quad (6.6)$$

$$T_{3k}^{(m,n)} - T_{3k}^{(m,n)} = 0 \quad \text{on } A_1 \times T \quad (6.7)$$

the macroscopic charge equation of electrostatics (5.17), the distribution of electric field (5.21) and that of strain by

$$S_{kl}^{(m,n)} = e_{kl}^{(m,n)} \quad (6.8)$$

the constitutive equations for the gross electric displacements (5.26) and those for the stress resultants in the form

$$T_{kl}^{(m,n)} = \sum_{p+q=0}^N I_{(m+p, n+q)} (C_{klrt} S_{rt}^{(p,q)} - C_{rkl} E_r^{(p,q)}) \quad (6.9)$$

the boundary conditions of surface charge (5.29) and (5.30), those of mechanical displacements (5.31) and those of electric potential (5.32), and the initial conditions (5.33)-(5.35). The linear governing

equations of quartz bar recover those by the author⁷ who has employed a semi-variational approach in his derivation.

Uniqueness of Solutions - The solutions of an initial mixed-boundary value problem defined by the one-dimensional linear governing equations of quartz bar are shown to be unique by means

of the logarithmic convexity arguments.⁴⁷ To establish this, as usual, the existence of two solutions arising from the same data $d_k^{(1)}$ and $d_k^{(2)}$ is supposed and the

difference solution $d_k (= d_k^{(1)} - d_k^{(2)})$ is considered.

The difference solution, that is, as before,

$u_k (= u_k^{(1)} - u_k^{(2)})$ and $\phi (= \phi^{(1)} - \phi^{(2)})$ evidently satisfies

the homogeneous parts of the governing equations by virtue of the linearity of these equations.

Accordingly, it suffices to show that the difference solution is trivial for the homogeneous governing equations in proving the uniqueness of solutions.

The treatment of uniqueness begins by defining the function $\mathcal{F}(t)$ by

$$\mathcal{F}(t) = \log F(t), \quad t \in T \quad (6.10a)$$

with

$$F(t) = \frac{1}{2} \int_A dx_3 \rho u_k u_k dA, \quad \tau_1 < t < \tau_2 \quad (6.10b)$$

$$F(t) = 0, \quad t \in [t_0, \tau_1] \text{ and } t \in [\tau_2, t_1] \quad (6.10c)$$

Here, Eq. (6.10c) clearly implies the uniqueness for all $t \in T$ but $t \in [\tau_1, \tau_2]$; $F(t)$ can be chosen, without loss of generality, as in this equation. Thus, only the interval $\tau = (\tau_1, \tau_2)$ is considered on which $F(t)$ is

positive by definition, and this function should satisfy the condition of the form

$$F^2 \ddot{\mathcal{F}} \equiv \dot{F}^2 - \ddot{\mathcal{F}} \geq 0, \quad \tau_1 < t < \tau_2 \quad (6.11)$$

for the convexity of $\mathcal{F}(t)$.

Now, the kinetic energy, K , the internal energy W and the total energy Ω per unit length of the quartz bar are calculated in the form

$$K = \frac{1}{2} \int_A \rho \dot{u}_k \dot{u}_k = \frac{1}{2} \sum_{m+n=0}^N \rho \dot{u}_k^{(m,n)} \dot{u}_k^{(m,n)} \quad (6.12)$$

$$\begin{aligned} W = \frac{1}{2} \int_A (\tau_{kl} e_{kl} + E_k D_k) dA = \frac{1}{2} [T_{3k}^{(m,n)} u_k^{(m,n)} \\ + (mT_{1k}^{(m-1, n)} + nT_{2k}^{(m, n-1)}) u_k^{(m,n)} \\ + D_3^{(m,n)} \phi^{(m,n)} + (mD_1^{(m-1, n)} + nD_2^{(m, n-1)}) \phi] \end{aligned} \quad (6.13)$$

$$\Omega = K + W \quad (6.14)$$

where the series expansions (4.1), the definitions (5.1), (5.7), (5.9) and (5.10) and the distributions (5.21) and (6.8) are used. Likewise, Eq. (6.10b) is evaluated with the result

$$F(t) = \frac{1}{2} \int_A \rho \sum_{m+n=0}^N u_k^{(m,n)} u_k^{(m,n)} dx_3 \quad (6.15)$$

in the interval τ . Then, time differentiations of this equation, by assuming the usual smoothness of functions, lead to

$$\dot{F}(t) = \int_A \rho \sum_{m+n=0}^N \dot{u}_k^{(m,n)} u_k^{(m,n)} dx_3 \quad (6.16)$$

and

$$\ddot{F}(t) = \int_A (2K + \rho \sum_{m+n=0}^N A_k^{(m,n)} u_k^{(m,n)}) dx_3 \quad (6.17)$$

in which Eqs. (5.7) and (6.12) are used. With the help of the homogeneous part of Eq. (6.5), Eq. (6.17) is expressed in the form

$$\begin{aligned} \ddot{F}(t) = \int_A [2K + \sum_{m+n=0}^N (T_{3k}^{(m,n)} - mT_{1k}^{(m-1, n)} \\ - nT_{2k}^{(m, n-1)} + P_k^{(m,n)}) u_k^{(m,n)}] dx_3 \end{aligned} \quad (6.18)$$

Then, on combining Eqs. (6.13) and (6.14) and integrating by parts, Eq. (6.18) takes the form

$$\ddot{F}(t) = -2W + \int_A 4K dx_3 + \Gamma + \chi \quad (6.19)$$

with

$$\Gamma = \sum_{m+n=0}^N (T_{3k}^{(m,n)} u_k^{(m,n)} + D_3^{(m,n)} \phi^{(m,n)})|_{x_3=0} \quad (6.20)$$

$$\chi = \int_A \sum_{m+n=0}^N (P_k^{(m,n)} u_k^{(m,n)} + D^{(m,n)} \phi^{(m,n)}) dx_3 \quad (6.21)$$

where Eq. (5.17) is taken into account. By the conservation of energy and the initial conditions (5.33)-(5.35), the total energy Ω is equal to zero. Besides, the boundary conditions (5.29)-(5.32), (6.6) and (6.7) render Γ and χ to zero, and then Eq. (6.19) becomes

$$\ddot{F}(t) = \int_A 4K dx_3 \quad (6.22)$$

In view of Eqs. (6.15), (6.16) and (6.22), one writes the right of Eq. (6.11) as

$$F^2 = \int_{\Omega} \left[\rho \sum_{m,n=0}^N [\dot{u}_k^{(m,n)} u_k^{(m,n)}] - [\dot{u}_k^{(m,n)} \dot{u}_k^{(m,n)}] \right] dx_3 - \int_{\Omega} \left[\rho \sum_{m,n=0}^N \dot{u}_k^{(m,n)} u_k^{(m,n)} \right] dx_3 \quad (6.23)$$

By virtue of Schwartz's inequality, one finds

$$F^2 \geq 0 \quad (6.24)$$

on the interval τ , and after integration, this implies

$$F(t) \leq [F(\tau_1)]^{\tau_2 - \tau_1 / \tau_2 - \tau_1} [F(\tau_2)]^{t - \tau_1 / \tau_2 - \tau_1} \quad \text{on } \tau \quad (6.25)$$

Due to the continuity of $F(t)$, $F(\tau_1)=0$, Eq. (6.25) shows that $F(t)=0$ for the interval τ as well, contrary to the initial hypothesis $F(t)>0$. Hence $F(t)=0$ for all $t \in T$, and the difference solution is trivial, that is, the uniqueness is established as in the case of polar rods.⁴⁸ A theorem of uniqueness is concluded as follows.

Theorem - Given a regular region of finite bar space $B+\partial B$ with its boundary surface $\partial B (=S_t U S_u = S_\sigma U S_\phi, S_t n S_u = S_\sigma n S_\phi = 0)$ in the Euclidean 3-Space Ξ , then there exists at most one set of twice continuously differentiable functions $u_k^{(m,n)}$ and $\phi^{(m,n)}$ in $B+\partial B$ at the time interval T , obeying Eqs. (5.17), (5.21), (5.26), (6.5), (6.8) and (6.9), and satisfying the boundary conditions (5.29)-(5.32), (6.6) and (6.7) and the initial conditions (5.33)-(5.35).

Quartz Bar With Initial Stresses - In the x_k -fixed system of Cartesian convected coordinates, consider the piezoelectric medium $B+\partial B$ with its boundary surface ∂B and closure \bar{B} . The medium is under initial stresses in its reference (initial) state which is considered to be self-equilibrating following loading in the natural state of medium. Then a small motion is superimposed upon the reference state. For this motion, the set of fundamental equations is consist of the stress equations of motion (2.1) and the boundary conditions of tractions (2.13) with the condition:^{49,50}

$$T_{kl} = \tau_{kr}^0 u_{l,r} \quad \text{in } \bar{B} \times T \quad (6.25)$$

or

$$(\tau_{kl} + \tau_{kr}^0 u_{l,r}),_k - \rho_0 a_l = 0 \quad \text{in } \bar{B} \times T \quad (6.26)$$

$$t_k^* - n_l (\tau_{lk} + \tau_{lr}^0 u_{k,r}) = 0 \quad \text{on } S_t \times T \quad (6.27)$$

the charge equation of electrostatics (2.3), the strain-mechanical displacement relations:

$$S_{kl} = e_{kl} = \frac{1}{2} (u_{k,l} + u_{l,k}) \quad \text{in } \bar{B} \times T \quad (6.28)$$

the electric field-electric potential relations (2.6), the constitutive relations

$$\tau_{kl} = C_{klmn} S_{mn} - C_{mkl} E_m \quad (6.29)$$

and Eq. (2.11), the boundary conditions of displacements, surface charge and electric potential (2.14)-(2.16) and the initial conditions (2.17) in the spatial (final) state. In the above equations; τ_{kl} , u_k , a_k and so on indicate small incremental quantities superimposed upon those of the reference state denoted

(0) such as $(\tau_{kl}^0, u_k^0, t_k^0)$. The incremental components of displacements u_k and the electric potential ϕ are represented by the series expansions (4.1). By paralleling to the derivation in Section 5, the macroscopic equations of thin quartz bar with initial stresses may be established by means of a variational principle²⁸ and the series expansions (4.1) as

$$\begin{aligned} & \tau_{3k}^{(m,n)} - m T_{3k}^{(m-1,n)} - n T_{2k}^{(m,n-1)} + p_k^{(m,n)} \\ & + N_k^{(m,n)} + R_k^{(m,n)} - \rho A_k^{(m,n)} = 0 \quad \text{on } L \times T \quad (6.30) \end{aligned}$$

$$T_k^{(m,n)} - (T_{3k}^{(m,n)} + N_{3k}^{(m,n)}) = 0 \quad \text{on } A_1 \times T \quad (6.31)$$

with the definitions (5.2) and (5.4)-(5.6) in terms of the incremental quantities, and alike.⁵¹

7. CONCLUSION

The main result presented herein is a set of one-dimensional, nonlinear electroelastic equations useful for analyzing wave propagation and vibrations in quartz bars. These governing equations of successively higher orders of approximation are deduced from the three-dimensional theory of piezoelectricity by a consistent method of reduction. That is, the variational principle (3.13) together with the series expansions (4.1) is used to derive the governing equations of quartz bar in which account is taken of only the elastic nonlinearities. The resulting equations incorporate as many higher order effects as deemed necessary in any case of interest by the proper truncation of the series expansions. Thus, the

customary use of matching coefficients⁶ is eliminated in a rational way. The nonlinear electroelastic equations describe all the higher order stretching, flexure and torsion of thin piezoelectric bars of uniform cross-section. Further, they are easily seen to reduce to the dynamic equations of bars by

Mindlin^{6,52}, Bleustein and Stanley⁵³, and the author^{7,8,51,54} as well as several authors mentioned by them.

The variational principle (3.13) is obtained from Hamilton's principle by modifying it through Friedrichs's transformation. As its Euler equations, the principle leads to all the fundamental equations of piezoelectricity but the initial conditions. By dropping out the nonlinear elastic terms, the variational principle can be specialized to contain

some of earlier variational principles.^{1,6,25,29,30,32,33,39,55,56}

The principle permits simultaneous approximation on all the field variables, and hence it is most frequently desirable and compulsory in selecting the trial functions of approximate direct

solutions.^{34,35,45} Further, special motions are pointed out, the linearized governing equations and the electroelastic equations in the presence of initial stresses are recorded for the quartz bar of uniform cross-section. The uniqueness is examined in solutions of the initial mixed-boundary value problem defined by the linearized governing equations, and the sufficient conditions for the uniqueness are enumerated by means of the logarithmic convexity arguments. It is worth noting that the uniqueness is established even though elasticities neither possess major symmetry (2.12) nor satisfy a definiteness condition of energies.^{47,56}

In closing, the results presented herein can be readily extended to the case in which the thermal effect⁵⁷⁻⁶⁰ and/or the mechanical effect of the electrode coating⁸ are taken into account. Likewise, for a piezoelectric bar with temperature-dependent properties,⁶¹ the nonlinear electroelastic equations of higher orders of approximation can be formulated. Further, it is worthwhile to conclude the paper that work⁴⁵ is now in progress for certain vibrations of

quartz bar, and it will be reported elsewhere.

ACKNOWLEDGEMENTS

The author wishes to extend his gratitude to Dr. Arthur Ballato and Dr. Julian J. Wu for their kind interest and encouragement. He also wishes to thank Ms. Ayse Inci for her excellent typing. The research reported herein has been sponsored in part by the United States Army through its European Research Office.

REFERENCES

1. M.C. Dökmeci, "Dynamic Applications of Piezoelectric Crystals" to appear in *The Shock and Vibration Digest*, Vol.18,1986
2. R.F. Milsom, D.T. Elliott, and M. Redwood, "Three-dimensional Mode-matching Theory of Rectangular Bar Resonators using Complex Wavenumbers," *Proc. 35th Ann. Freq. Control Symp.*, IEEE, New York, 1981, pp.174-186.
3. R.F. Milsom, D.T. Elliott, S. Terry-wood, and M. Redwood, "Analysis and Design of Coupled-Mode Miniature Bar Resonators and Monolithic Filters," *IEEE SU-30*, No.3, 1983, pp.140-255.
4. H.F. Tiersten and A. Ballato, "Nonlinear Extensional Vibrations of Quartz Rods," *J. Acoust. Soc. Am.*, Vol.73, No.6,1983, pp.2022-2033.
5. A.E. Green and P.M. Naghdi, "Electromagnetic Effects in the Theory of Rods," *Phil. Trans. R. Soc. Lond. A* Vol.314,1985, pp.311-352.
6. R.D. Mindlin, *Theory of Beams and Plates*, Lecture notes at Columbia University, New York, 1968.
7. M.C. Dökmeci, "A Theory of High-frequency Vibrations of Piezoelectric Crystal Bars," *Int. J. Solids Structures*, Vol.10, No.4,1974, pp.401-409.
8. M.C. Dökmeci, "High Frequency Vibrations of Piezoelectric Bars, Coated with Electrodes," *Proc. 12th Annual Soc. Engng. Science*, The Univ. Texas at Austin, 1975, pp.433-442.
9. M.C. Dökmeci, "A Piezoelectric Model for Vibrations of Bones," *Development in Mechanics*, Vol.13, Ohio State University, Columbus, Ohio, 1985, pp.523-524.
10. A.A. Bondarenko, G.V. Kutsenko, and A.F. Ulitko, "Amplitudes and Phases of Longitudinal Vibrations of Piezoceramic Rods with Account of Variable Mechanical Quality Factor," *Sov. Appl. Mech.*, Vol.16, No.11,1981, pp.1001-1004.
11. A. Das and A. Ray, "Forced Transient Motion of a Piezoelectric Bar and Its Voltage Response," *Indian J. Technology*, Vol.18,1980, pp.349-353.
12. O. Yu. Zharii, "Discharge of a Piezoceramic Rod in Shock Loading," *Sov. Appl. Mech.*, Vol.17, No.3,1981, pp.284-288.
13. E.P. Eer Nisse, "Electrode Stress Effects for Length-Extensional and Flexural Resonant Vibrations of Quartz Bars," *Proc. 38th Ann. Freq. Control Symp.*, IEEE, New York, 1984, pp.126-131.
14. I.A. Sazonov, "Waveguide Vibration Isolator for Arbitrarily Polarized Flexural Modes," *Sov. Phys. Acoust.*, Vol.28, No.4,1983, pp.321-323.
15. K.K. Kundu, "Mechanical Disturbances in a Piezoelectric Bar with Heat Influx at One End," *Indian J. Technology*, Vol.19,1981, pp.504-506.
16. E. Bassiouny and A.F. Ghaleb, "A One-Dimensional Problem in the Generalized Theory of Thermoelastopiezoelectricity," *The Mechanical Behavior of Electromagnetic Solid Continua*, North-Holland, 1984, pp.79-84.
17. T. Ikeda, "Depolarizing-field Effect in Piezoelectric Resonators," *Ferroelectrics*, Vol.43, 1982, pp.3-15.
18. N.F. Ivina and B.A. Kasatkin, "Numerical Analysis of the Dispersion Relations for the Normal Modes of a Piezoelectrically Active Waveguide," *Sov. Phys. Acoust.*, Vol.28, No.4,1982, pp.303-305.
19. A.A. Bondarenko, N.I. Karas, and A.F. Ulitko, "Methods for Determining the Vibrational Dissipation Characteristics of Piezoceramic Structural Elements," *Sov. Appl. Mech.*, Vol.18, No.2,1982, pp.175-179.
20. I.M. Zadirienko and A.G. Kuz'menko, "Synthesizing the Electrical Excitation of an Asymmetrically Radiating Sectional Bar Transducer," *Sov. Phys. Acoust.*, Vol.29, No.6,1983, pp.453-455.
21. A.N. Solov'ev, "Influence of the Electroded Zone on the Natural Frequency of a Piezoceramic Body of Rectangular Cross Section," *Sov. Appl. Mech.*, Vol.20, No.9,1985, pp.837-843.
22. H.F. Tiersten, "Nonlinear Electroelastic Equations Cubic in the Small Field Variables," *J. Acoust. Soc. Am.*, Vol.57, No.3,1975, pp.660-666.
23. H.F. Tiersten, "Electroelastic Interactions and the Piezoelectric Equations," *J. Acoust. Soc. Am.*, Vol.70, No.6,1981, pp.1567-1576.
24. G. Venkataraman, L.A. Feldkamp, and V.C. Sahni, *Dynamics of Perfect Crystals*, M.I.T. Press, Cambridge, 1975.
25. H.F. Tiersten, *Linear Piezoelectric Plate Vibrations*, Plenum Press, New York, 1969.
26. P.P. Silvester and R.L. Ferrari, *Finite Elements for Electrical Engineers*, Cambridge Univ. Press, London, 1983.
27. P. Hammond, *Energy Methods in Electromagnetism*, Clarendon Press, Oxford, 1981.
28. M.C. Dökmeci, "Vibrations of Piezoelectric Discs under Initial Stresses," *Proc. 39th Ann. Freq. Control Symp.*, IEEE, New York, 1984, pp.431-435.
29. M.C. Dökmeci and N. Sarıgül, "Certain Integral and Differential Types of Variational Principles of Mathematical Physics," paper at the Int. Congress of Mathematicians, Univ. California, Berkeley, August 1986.
30. M.C. Dökmeci, "The Principle of Virtual Work Applied to Nonlinear Piezoelectric Continuum and Some Associated Variational Principles," paper at the 23rd Annual Meeting of the Soc. Engng. Science, State Univ. New York, Buffalo, August 1986.
31. M.V.K. Chari and P.P. Silvester (Eds.), *Finite Elements in Electrical and Magnetic Field Problems*, John Wiley, New York, 1980.

32. M.C. Dökmeci, "Variational Principles in Piezoelectricity," L. Al Nuovo Cimento, Vol.7, No.11, 1983, pp.449-454.
33. M.C. Dökmeci, "Variational Theorems Associated with Hamilton's Principle," Proc. 4th Can. Cong. Appl. Mechs., École Polytechnique, Montréal, 1973, pp.515-516.
34. L.V. Kantorovich and V.I. Krylov, Approximate Methods of Higher Analysis, Interscience Publ., New York, 1964.
35. L. Collatz, The Numerical Treatment of Differential Equations, Springer-Verlag, Berlin, 1960.
36. H.F. Tiersten, "Natural Boundary and Initial Conditions from a Modification of Hamilton's Principle," J. Math. Phys., Vol.9, No.9, 1968, pp. 1445-1451.
37. M.E. Gurtin, "The Linear Theory of Elasticity," Handbuch der Physik, Vol.VIa/2, Springer-Verlag, Berlin, 1972, pp.1-295.
38. R. Courant and D. Hilbert, Methods of Mathematical Physics, Vol.1, Interscience Publ., New York, 1965.
39. M.C. Dökmeci, "Recent Advances: Vibrations of Piezoelectric Crystals," Int. J. Engng. Sci., Vol.18, No.3, 1980, pp.431-448.
40. M.C. Dökmeci, "Techniques in Reducing Equations of Mathematical Physics," paper at Int. Congress of Mathematicians, Helsinki (1978).
41. M.C. Dökmeci, Vibration and Stability Analysis of Laminated Composites, A.E.K. Tech. R.2940-3-77, Ankara (1979).
42. Z. Nikodem and P.C.Y. Lee, "Approximate Theory of Vibration of Crystal Plates at High Frequencies," Int. J. Solids Structures, Vol.10, 1974, pp.177-196.
43. S. Syngellakis and P.C.Y. Lee, "An Approximate Theory for the High Frequency Vibrations on Piezoelectric Crystal Plates," Proc. 30th Ann. Symp. on Freq. Control, U.S. Army Electronics Command, Fort Monmouth, New Jersey, 1976, pp.184-190.
44. IEEE Standard on Piezoelectricity, IEEE, New York, 1979.
45. M.C. Dökmeci and N. SrgulD, "Low-Frequency Extensional Vibrations of Quartz Rods," forthcoming.
46. Y.C. Fung, Foundations of Solid Mechanics, Prentice Hall, Englewood Cliffs, New Jersey, 1965.
47. R.J. Knops and L.E. Payne, Uniqueness Theorems in Linear Elasticity, Springer-Verlag, Berlin, 1972.
48. M.C. Dökmeci, "An Isothermal Theory of Anisotropic Rods," J. Engng. Maths., Vol.9, 1975, pp.311-322.
49. V.V. Bolotin, Nonconservative Problems of the Theory of Elastic Stability, Pergamon Press, London, 1963.
50. A.L. Nalamwar and M. Epstein, "Surface Acoustic Waves in Strained Media," J. Appl. Phys., Vol.47, 1976, pp.43-48.
51. M.C. Dökmeci, "Variational Theorems for Superimposed Motions in Elasticity, with Application to Beams," Recent Advances in Engineering Science, Vol.11, NASA CP-2001, 1976, pp.481-490.
52. R.D. Mindlin, "Low Frequency Vibrations of Elastic Bars," Int. J. Solids Structures, Vol.12, 1976, pp.27-49.
53. J.L. Bleustein and R.M. Stanley, "A Dynamical Theory of Torsion," Int. J. Solids Structures, Vol.6, 1970, pp.569-586.
54. M.C. Dökmeci, "A General Theory of Elastic Beams," Int. J. Solids Structures, Vol.8, 1972, pp.1205-1222.
55. M.C. Dökmeci, "Dynamic Variational Principles for Discontinuous Elastic Fields," J. Ship Research, Vol.23, 1979, pp.115-122.
56. C.E. Beevers and R.E. Craine, "On the Thermodynamics and Stability of Deformable Dielectrics," Appl. Scientific Research, Vol.37, 1981, pp.241-256.
57. C.E. Beevers and R.E. Craine, "Wave Propagation in a Thermoelastic Dielectric," J. Mecanique théorique et appliquée, Vol.4, No.2, 1985, pp.159-174.
58. M.C. Dökmeci, "Theory of Vibrations of Coated, Thermopiezoelectric Laminae," J. Math. Phys., Vol. 19, No.1, 1978, pp.109-126.
59. R.D. Mindlin, "Equations of High Frequency Vibrations of Thermopiezoelectric Crystal Plates," Int. J. Solids Structures, Vol.10, 1974, pp.625-637.
60. Y. Mengi and N. Akkas, "A New Higher Order Dynamic Theory for Thermoelastic Bars. I: General Theory," J. Acoust. Soc. Am., Vol.73, No.6, 1983, pp.1918-1922.
61. M.C. Dökmeci and N. Sarıgül, "Vibrations of Elastically Supported Thermoelastic Cylindrical Shells with Temperature-Dependent Properties," paper at the 23rd Annual Meeting of the Soc. Engng. Science, State Univ. New York, Buffalo, August 1986.

THREE-DIMENSIONAL FINITE ELEMENT SOLUTION OF THE LAGRANGEAN EQUATIONS
FOR THE FREQUENCY-TEMPERATURE BEHAVIOR OF Y-CUT AND NT-CUT BARS.

Y.K. Yong
Dept. of Civil/Environmental Engineering
Rutgers University
P.O.Box 909
Piscataway, N.J. 08854

SUMMARY

A three-dimensional finite element matrix equation was formulated, using variational principles, from the field equations of incremental motion superposed on homogenous thermal strain. The field equations were derived from the nonlinear field equations of thermoelasticity in Lagrangean formulation. Since the equations were referred to a fixed reference frame, the element nodal coordinates and mass matrix were not updated with changes in temperature. Only the stiffness matrix must be updated. An anisotropic plate equation was derived and it was observed that only one term β_{22} of the thermal expansion coefficient β_{ij} appears in the equation. Hence, for low-frequency flexural vibrations, the tensor β_{ij} was assumed to be equal to $\beta_{22} \mathbf{I}$ in the three-dimensional finite element analysis. A hexahedral element with eight nodes and three degrees of freedom per node was used. Guyan reduction scheme was employed to reduce the mass and stiffness matrices. Either the Householder or the full subspace iteration method was used to extract the eigenvalues for frequency calculations. Results using the finite element method were compared with the analytical and experimental results for a Y-cut plate, NT-cut bars and tuning forks.

INTRODUCTION

For low-frequency flexural vibrations of quartz resonators, the classical beam equation is commonly used due to the simplicity of its expression for resonance frequency. This equation, however, cannot take into account the change in the frequency-temperature behavior due to a change in the width to thickness ratio. The frequency-temperature behavior of the rotated X-cut tuning forks of Nakazawa, Nakamura and Miyashita¹ were shown empirically to be a function of both the length to thickness and width to thickness ratios. In a previous paper², closed form solution of a set of coupled one-dimensional plate equations for the frequency-temperature behavior of Y-cut plates was shown to be dependent upon both the length to thickness and width to thickness ratios. Some recent

works^{3,4,5} employed the finite element method in the study of the quartz tuning forks using a three dimensional hexahedral element.

Lee and Yong⁶ derived the three-dimensional incremental field equations for small vibrations superposed on homogenous thermal strains from the nonlinear field equations of thermoelasticity in Lagrangian formulation, and calculated the three orders of the temperature derivatives of elastic constants using the experimental results given by Bechmann, Ballato and Lukaszek. In this paper, the three-dimensional equations were casted in a variational form using Hamilton's principle from which the finite element matrix equation was derived. Since the field equations were referred to a fixed reference frame, a temperature change did not affect the consistent mass matrix and the element nodal coordinates. Only the stiffness matrix was updated due to changes in the temperature related material properties. This is in contrast with the theory based on linear elasticity¹³ with the assumption that the material parameters such as the elastic constants, density and plate dimensions are a cubic polynomial function of the change in temperature. The reference state is not fixed; hence its finite element implementation would involve not only the updating of the stiffness matrix but also the element nodal coordinates and the consistent mass matrix.

An anisotropic plate equation was derived from a set of two-dimensional incremental plate to show that one component of the thermal expansion coefficients tensor β_{ij} is predominant, namely, β_{22} . Hence, when the three-dimensional finite element model was employed in the study of low-frequency flexural vibrations of crystal bars, the tensor β_{ij} was assumed to be equal to $\beta_{22} \delta_{ij}$.

This assumption simplifies the three-dimensional finite element calculations. An anisotropic hexahedral element with eight nodes and three translational degrees of freedom per node was used. Depending upon the boundary conditions, either the Householder method or the full subspace iteration method was used to extract the eigenvalues for frequency-temperature calculations. The

Householder method was used when the stiffness matrix was not positive definite as in the case when the resonator was subjected to only traction-free conditions. Master degrees of freedom obtained by static condensation of the stiffness and mass matrices were chosen along the longitudinal centerline of the bar. Results for the frequency-temperature behavior of a Y-cut plate and NT-cut bars and tuning forks were compared with the available experimental results by Singer and Paros and Nakazawa, Nakamura and Miyashita¹. Good overall agreement was found.

FINITE ELEMENT FORMULATION OF THE THREE-DIMENSIONAL INCREMENTAL EQUATIONS OF MOTION SUPERPOSED ON HOMOGENEOUS THERMAL STRAINS.

In this section, the matrix notation commonly used in finite element formulation will be used. For a body occupying a region V with bounding surface S , the variational principle of elasticity⁹ states that

$$\delta \int_{t_0}^{t_1} (K - V) dt + \int_{t_0}^{t_1} \delta W dt = 0 \quad (1)$$

where t is time, K and V the total kinetic and strain energies in V , and δW the work done by the surface traction over S through varied displacement.

To apply the variational principle to an elastic body moving from the initial state at temperature, T , different from T_0 , the reference temperature, to the final state of incremental motion superposed on the initial state⁶, we let

$$K = \int_V \frac{1}{2} \rho \dot{\underline{u}}^T \underline{\dot{u}} dV, \quad (2)$$

$$V = \int_V \underline{e}^T \left(\underline{T} + \frac{1}{2} \underline{t} \right) dS,$$

$$\text{and } \delta W = \int_S \delta \underline{u}^T \underline{p} dS.$$

where the work done by the body force has been omitted for simplicity. The vectors \underline{T} , \underline{u} , \underline{e} , \underline{t} and \underline{p} are respectively, the initial stress and the incremental displacement, strain, stress, and surface traction vectors which consist of the following components:

$$\underline{T} = [T_{11}, T_{22}, T_{33}, T_{23}, T_{13}, T_{12}]^T \quad (3)$$

$$\underline{u} = [u_1, u_2, u_3]^T$$

$$\underline{e} = [e_{11}, e_{22}, e_{33}, 2e_{23}, 2e_{13}, 2e_{12}]^T$$

$$\underline{t} = [t_{11}, t_{22}, t_{33}, t_{23}, t_{13}, t_{12}]^T$$

$$\underline{p} = [p_1, p_2, p_3]^T$$

The strain-displacement-temperature and stress-strain-temperature relations are given as⁶

$$\underline{e} = \underline{D} \underline{b} \underline{u} \quad \text{and} \quad \underline{t} = \underline{D} \underline{e} \quad (4)$$

where d and b are the linear differential operator and the thermal expansion coefficient tensor, respectively, equal to:

$$\begin{bmatrix} \partial/\partial x_1 & 0 & 0 \\ 0 & \partial/\partial x_2 & 0 \\ 0 & 0 & \partial/\partial x_3 \\ & \partial/\partial x_3 & \partial/\partial x_2 \\ \partial/\partial x_3 & 0 & \partial/\partial x_1 \\ \partial/\partial x_2 & \partial/\partial x_1 & \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{12} & \beta_{22} & \beta_{23} \\ \beta_{13} & \beta_{23} & \beta_{33} \end{bmatrix} \quad (5)$$

The material tensors \underline{D} and \underline{b} are functions of temperature change, $\theta = (T - T_0)$:

$$D_{pq} = C_{pq} + D_{pq}^{(1)} \cdot \theta + D_{pq}^{(2)} \cdot \theta^2 + D_{pq}^{(3)} \cdot \theta^3 \quad (6)$$

$$\beta_{ij} = \delta_{ij} + \alpha_{ij}^{(1)} \cdot \theta + \alpha_{ij}^{(2)} \cdot \theta^2 + \alpha_{ij}^{(3)} \cdot \theta^3$$

δ_{ij} is the Kronecker delta and the values⁶ of C_{pq} , $D_{pq}^{(n)}$ and $\alpha_{ij}^{(n)}$ for the Y-cut at 25°C are given in the appendix.

By the use of integration by parts and the condition that $\delta \underline{u}_i$ vanishes at t_0 and t_1 , we obtain

$$\delta \int_{t_0}^{t_1} K dt = - \int_{t_0}^{t_1} dt \int_V \delta \underline{u}^T \rho \dot{\underline{u}} dV \quad (7)$$

By setting $\underline{T} = 0$ for zero initial stresses and inserting the stress-strain-temperature relations (4)₂, the variational strain energy becomes

$$\delta V = \int_V \delta \underline{e}^T \underline{t} dV = \int_V \delta \underline{e}^T \underline{D} \underline{e} dV \quad (8)$$

In the finite element method, the generic incremental displacement \underline{u} is related to the element nodal displacements \underline{z} by certain assumed interpolating displacement shape functions¹⁰:

$$\underline{u} = \underline{f} \underline{z} \quad (9)$$

where $\underline{z} = (z_1, z_2, \dots, z_{N_e})^T$, N_e = number of

degrees of freedom of the element, and \underline{f} denotes a rectangular matrix containing the functions that make \underline{u} completely dependent upon \underline{z} . If an eight node hexahedron element with three degrees of freedom at each node is used, the matrix sizes of \underline{z} and \underline{f} are, respectively, 24x1 and 3x24.

By a simple algebraic manipulation, using expressions from (2)₃, (4)₁, (7), (8) and (9), eqn (1) may be rewritten as:

$$\int_{t_0}^{t_1} dt \delta \underline{z}^T \left[\int_V \left[-\rho \underline{f}^T \underline{\ddot{z}} - \underline{f}^T (\underline{d} \underline{b}) \underline{T} \underline{D} (\underline{d} \underline{b}) \underline{f} \underline{z} \right] dV + \int_S \underline{f}^T \underline{p} dS \right] = 0 \quad (10)$$

For independent variations $\delta \underline{z}^T$, eqn. (10) yields the element stiffness equation:

$$\underline{M} \underline{\ddot{z}} + \underline{K} \underline{z} = \underline{F} \quad (11)$$

where $\underline{M} = \int_V \rho \underline{f}^T \underline{f} dV$ Consistent Mass

$\underline{K} = \int_V (\underline{d} \underline{b} \underline{f})^T \underline{D} (\underline{d} \underline{b} \underline{f}) dV$ Stiffness

$\underline{F} = \int_S \underline{f}^T \underline{p} dS$ Surface Traction

For free vibrations, \underline{F} is set to zero and assuming harmonic motion $\underline{z} = \underline{A} e^{i\omega t}$, eqn (11) yields

$$(\underline{K} - \omega_i^2 \underline{M}) \underline{A}_i = 0 \quad (12)$$

where ω_i ($i = 1, 2, \dots, N$) are the natural frequencies and \underline{A}_i the corresponding modes.

Since the governing field equations are referred to a fixed reference frame, the density and the nodal coordinates of the elements do not change with temperature. Hence, the mass matrix need not be updated with the temperature change. Only the material matrices \underline{b} and \underline{D} in the stiffness matrix \underline{K} is updated. This is in contrast with the theory¹³ based on linear elasticity where the reference frame is not fixed which necessitate the updating of the element nodal coordinates, mass matrix and stiffness matrix.

ANISOTROPIC INCREMENTAL PLATE EQUATION SUPERPOSED ON HOMOGENEOUS THERMAL STRAIN.

Since the study involved low-frequency flexural vibrations, the form of the anisotropic plate is investigated. The two-dimensional incremental equations of motion for small amplitude waves superposed on homogenous thermal strains⁶ were given as follows:

N^{th} Order Stress-Temperature Equations of Motion

$$\beta_{ik} t_{kj,j}^{(n)} - n \beta_{ik} t_{k2}^{(n-1)} + \beta_{ik} F_k^{(n)} = \rho \sum_{m=0}^{\infty} A_{mn} \ddot{u}_i^{(m)}$$

where $t_{kj}^{(n)} = \int_{-b}^b x_2^n t_{kj} dx_2$

$$F_k^{(n)} = \left[x_2^n t_{k2} \right]_{-b}^b$$

and $A_{mn} = \frac{2b^{m+n+1}}{m+n+1}$ when $(m+n)$ is even

$= 0$ when $(m+n)$ is odd

(13)

The natural boundary conditions are prescribed along a curve C bounding the edges of the plate:

$$p_i^{(n)} = \beta_{ik} n_j t_{kj}^{(n)} \quad \text{on } C \quad (14)$$

n_j are the outward normal components on C .

N^{th} Order Strain-Displacement-Temperature Relations.

$$e_{ij}^{(n)} = \frac{1}{2} \left[\beta_{kj} u_{k,i}^{(n)} + \beta_{ki} u_{k,j}^{(n)} + (n+1) \left[\delta_{2i} \beta_{kj} u_k^{(n+1)} + \delta_{2j} \beta_{ki} u_k^{(n+1)} \right] \right] \quad (15)$$

N^{th} Order Stress-Strain-Temperature Relations.

$$t_{ij}^{(n)} = \sum_{m=0}^{\infty} A_{mn} D_{ijkl} e_{kl}^{(m)} \quad (16)$$

A procedure proposed by Mindlin⁷ is used to reduce the two-dimensional equations to a single anisotropic flexural equation. The first order plate equations are extracted from the infinite series of equations by setting $t_{ij}^{(n)}$ and $e_{ij}^{(n)}$ equal to zero for $n > 1$ so as to obtain:

$$\beta_{ik} t_{kj,j}^{(0)} = 2b \rho \ddot{u}_i^{(0)} \quad (17)$$

$$\beta_{ik} t_{kj,j}^{(1)} - \beta_{ik} t_{k2}^{(0)} = \frac{2b^3}{3} \rho \ddot{u}_i^{(1)}$$

Specify

$$1) p_i^{(0)} = \beta_{ik} n_j t_{kj}^{(0)} \quad \text{or } u_i^{(0)} \quad \text{on } C \quad (18)$$

$$2) p_i^{(1)} = \beta_{ik} n_j t_{kj}^{(1)} \quad \text{or } u_i^{(1)} \quad \text{on } C$$

$$t_{ij}^{(0)} = 2b\kappa_{(ij)}\kappa_{kl}D_{ijkl}e_{kl}^{(0)} \quad (19)$$

$$t_{ij}^{(1)} = \frac{2b^3}{3}D_{ijkl}e_{kl}^{(1)}$$

$$e_{ij}^{(0)} = \frac{1}{2} \left[\beta_{kj}u_{k,i}^{(0)} + \beta_{ki}u_{k,j}^{(0)} + \left[\delta_{2i}\beta_{kj}u_k^{(1)} + \delta_{2j}\beta_{ki}u_k^{(1)} \right] \right] \quad (20)$$

$$e_{ij}^{(1)} = \frac{1}{2} \left[\beta_{kj}u_{k,i}^{(1)} + \beta_{ki}u_{k,j}^{(1)} \right]$$

where $F_1^{(0)}$ and $F_1^{(1)}$ were set to zero in the equations of motion for the traction free conditions on the plate major surface. κ_{ij} ($=\kappa_{ji}$) are the shear correction factors introduced in reference 8 for plates initially subjected to static, homogenous strains and the indices in parenthesis are suspended from summation.

$$\kappa_{12}^2 = \kappa_{22}^2 = \kappa_{23}^2 = \kappa^2 = \frac{\pi^2}{12} \quad (21)$$

$$\kappa_{11}^2 = \kappa_{33}^2 = \kappa_{13}^2 = 1$$

For thickness-shear and thickness-flexure vibrations, we may set

$$t_{22}^{(0)} = t_{11}^{(0)} = t_{33}^{(0)} = t_{13}^{(0)} = 0 \quad (22)$$

$$t_{22}^{(1)} = t_{23}^{(1)} = t_{12}^{(1)} = 0$$

To allow for the free development of the corresponding strains:

$$e_{22}^{(0)}, e_{11}^{(0)}, e_{33}^{(0)}, e_{13}^{(0)}, e_{22}^{(1)}, e_{23}^{(1)} \text{ and } e_{12}^{(1)} \quad (23)$$

the stress-strain relations (19) are first written in terms of elastic compliances S_{pq} instead of the elastic stiffnesses D_{pq} :

$$2b\kappa^2 e_4^{(0)} = S_{44}t_4^{(0)} + S_{46}t_6^{(0)} \quad (24)$$

$$2b\kappa^2 e_6^{(0)} = S_{46}t_4^{(0)} + S_{66}t_6^{(0)}$$

$$\frac{2b^3}{3}e_1^{(1)} = S_{11}t_1^{(1)} + S_{13}t_3^{(1)} + S_{15}t_5^{(1)} \quad (25)$$

$$\frac{2b^3}{3}e_3^{(1)} = S_{13}t_1^{(1)} + S_{33}t_3^{(1)} + S_{35}t_5^{(1)}$$

$$\frac{2b^3}{3}e_5^{(1)} = S_{15}t_1^{(1)} + S_{35}t_3^{(1)} + S_{55}t_5^{(1)}$$

where the abbreviated indicial notation was used. That is,

$$t_{ij}^{(n)} = t_p^{(n)} \quad i=j, \quad p=1, 2, 3$$

$$i \neq j, \quad p=4, 5, 6$$

$$e_{ij}^{(n)} = e_p^{(n)} \quad i=j, \quad p=1, 2, 3$$

$$= \frac{1}{2}e_p^{(n)} \quad i \neq j, \quad p=4, 5, 6$$

Solving for the stress components in terms of strains yields

$$t_4^{(0)} = 2b\kappa^2(\bar{D}_{44}e_4^{(0)} + \bar{D}_{46}e_6^{(0)}) \quad (26)$$

$$t_6^{(0)} = 2b\kappa^2(\bar{D}_{46}e_4^{(0)} + \bar{D}_{66}e_6^{(0)})$$

$$t_1^{(1)} = \frac{2b^3}{3}(\bar{D}_{11}e_1^{(1)} + \bar{D}_{13}e_3^{(1)} + \bar{D}_{15}e_5^{(1)}) \quad (27)$$

$$t_3^{(1)} = \frac{2b^3}{3}(\bar{D}_{13}e_1^{(1)} + \bar{D}_{33}e_3^{(1)} + \bar{D}_{35}e_5^{(1)})$$

$$t_5^{(1)} = \frac{2b^3}{3}(\bar{D}_{15}e_1^{(1)} + \bar{D}_{35}e_3^{(1)} + \bar{D}_{55}e_5^{(1)})$$

$$\text{where } \bar{D}_{44} = \frac{S_{66}}{\Delta_1}, \bar{D}_{46} = -\frac{S_{46}}{\Delta_1}, \bar{D}_{66} = \frac{S_{44}}{\Delta_1} \text{ and}$$

$$\Delta_1 = S_{44}S_{66} - S_{46}^2$$

$$\text{and } \bar{D}_{11} = \frac{[S_{33}S_{55} - S_{35}^2]}{\Delta_2}; \bar{D}_{13} = \frac{[S_{15}S_{35} - S_{13}S_{55}]}{\Delta_2} \quad (28)$$

$$\bar{D}_{33} = \frac{[S_{11}S_{55} - S_{15}^2]}{\Delta_2}; \bar{D}_{15} = \frac{[S_{13}S_{35} - S_{15}S_{33}]}{\Delta_2}$$

$$\bar{D}_{55} = \frac{[S_{11}S_{33} - S_{13}^2]}{\Delta_2}; \bar{D}_{35} = \frac{[S_{13}S_{15} - S_{11}S_{35}]}{\Delta_2}$$

$$\Delta_2 = S_{11}S_{33}S_{55} + 2S_{13}S_{15}S_{35} - S_{11}S_{35}^2 - S_{33}S_{15}^2 - S_{55}S_{13}^2$$

The extensional displacements $u_1^{(0)}, u_3^{(0)}$

and $u_2^{(1)}$ were accommodated in the strains of eqn. (23) and may be neglected in all subsequent developments. Hence, we are left with three stress equations of motion,

$$\beta_{22}t_{12,1}^{(0)} + \beta_{22}t_{23,3}^{(0)} - 2b\rho \ddot{u}_2^{(0)} \quad (29)$$

$$\beta_{1k}t_{1k,1}^{(1)} + \beta_{1k}t_{3k,3}^{(1)} - \beta_{1k}t_{k2}^{(0)} = 2b\rho \ddot{u}_1^{(1)}$$

$$\beta_{3k}t_{1k,1}^{(1)} + \beta_{3k}t_{3k,3}^{(1)} - \beta_{3k}t_{k2}^{(0)} = 2b\rho \ddot{u}_3^{(1)}$$

In order to obtain a single flexural equation similar in form to Cauchy's anisotropic plate equation, the transverse shear strains $e_{23}^{(0)}$ and $e_{12}^{(0)}$ are set to zero, which is equivalent to the assumption that a line remains perpendicular to the neutral surface of the plate during deformation:

$$e_{23}^{(0)} = \frac{1}{2}(\beta_{22}u_{2,3}^{(0)} + \beta_{r3}u_r^{(1)}) = 0$$

$$e_{12}^{(0)} = \frac{1}{2}(\beta_{22}u_{2,1}^{(0)} + \beta_{r1}u_r^{(1)}) = 0$$

$$\text{Therefore, } \beta_{rs}u_r^{(1)} = -\beta_{22}u_{2,s}^{(0)} \quad (30)$$

Implicit in the above assumption is that the elastic stiffnesses \bar{D}_{2323} and \bar{D}_{1212} are infinitely large (or stiff). The corresponding stresses $t_{23}^{(0)}$ and $t_{12}^{(0)}$ becomes indefinite.

Equations (29)₂ and (29)₃ are used to yield

definite values of $t_{23}^{(0)}$ and $t_{12}^{(0)}$ by

neglecting the rotary inertias $\frac{2b^3}{3}\rho \ddot{u}_1^{(1)}$ and $\frac{2b^3}{3}\rho \ddot{u}_1^{(1)}$ to yield

$$\beta_{1k} t_{k2}^{(0)} = \beta_{1k} t_{1k,1}^{(1)} + \beta_{1k} t_{3k,3}^{(1)} \quad (31)$$

$$\beta_{3k} t_{k2}^{(0)} = \beta_{3k} t_{1k,1}^{(1)} + \beta_{3k} t_{3k,3}^{(1)}$$

$$\text{or } t_{k2}^{(0)} = t_{1k,1}^{(1)} + t_{3k,3}^{(1)} \quad k = 1, 3 \quad (32)$$

since β_{1k} and β_{3k} are common in eqns. (31)

The stress-displacement relations are obtained by substituting the strains in eqns. (27) with the strain-displacement relations (20)₂. When the displacements in the stress-displacement relations are substituted with the expression given in (30), we obtain the following stress-displacement relations in terms of $u_2^{(0)}$

$$\begin{aligned} t_1^{(1)} &= -\frac{2b^3}{3} \left[\bar{D}_{11} \beta_2 u_{2,11}^{(0)} + \bar{D}_{13} \beta_2 u_{2,33}^{(0)} + 2\bar{D}_{15} \beta_2 u_{2,13}^{(0)} \right] \\ t_3^{(1)} &= -\frac{2b^3}{3} \left[\bar{D}_{13} \beta_2 u_{2,11}^{(0)} + \bar{D}_{33} \beta_2 u_{2,33}^{(0)} + 2\bar{D}_{35} \beta_2 u_{2,13}^{(0)} \right] \\ t_5^{(1)} &= -\frac{2b^3}{3} \left[\bar{D}_{15} \beta_2 u_{2,11}^{(0)} + \bar{D}_{35} \beta_2 u_{2,33}^{(0)} + 2\bar{D}_{55} \beta_2 u_{2,13}^{(0)} \right] \end{aligned} \quad (33)$$

where $\beta_2 = \beta_{22}$.

Substituting the right hand side of (32) with the above eqns. (33) yields:

$$\begin{aligned} t_4^{(0)} &= -\frac{2b^3}{3} \left[\beta_2 \bar{D}_{15} u_{2,11}^{(0)} + (\beta_2 \bar{D}_{13} + 2\beta_2 \bar{D}_{55}) u_{2,113}^{(0)} \right. \\ &\quad \left. + 3\beta_2 \bar{D}_{35} u_{2,133}^{(0)} + \beta_2 \bar{D}_{33} u_{2,333}^{(0)} \right] \\ t_6^{(0)} &= -\frac{2b^3}{3} \left[\beta_2 \bar{D}_{11} u_{2,11}^{(0)} + (\beta_2 \bar{D}_{13} + 2\beta_2 \bar{D}_{55}) u_{2,133}^{(0)} \right. \\ &\quad \left. + 3\beta_2 \bar{D}_{15} u_{2,113}^{(0)} + \beta_2 \bar{D}_{35} u_{2,333}^{(0)} \right] \end{aligned} \quad (34)$$

Finally, a single flexural displacement equation of motion is obtained by substituting eqns. (34) into the left hand side of (29)₁:

$$\begin{aligned} \beta_2^2 \bar{D}_{11} u_{2,1111}^{(0)} + 4\beta_2^2 \bar{D}_{15} u_{2,1113}^{(0)} + \\ 2(\beta_2^2 \bar{D}_{13} + 2\beta_2^2 \bar{D}_{55}) u_{2,1133}^{(0)} + 4\beta_2^2 \bar{D}_{35} u_{2,1333}^{(0)} + \\ \beta_2^2 \bar{D}_{33} u_{2,3333}^{(0)} = -\frac{3\rho}{b^2} \ddot{u}_2^{(0)} \end{aligned} \quad (35)$$

We observe that only one element, β_2 , of the thermal expansion coefficient tensor, β_{ij} , appears in the above governing equation. In the three-dimensional finite analysis of the frequency-temperature behavior of quartz resonators, the thermal expansion coefficient, \underline{b} , in eqn. (11)₃ is assumed to be scalar valued:

$$\underline{b} = \beta_2 \underline{I} \quad (36)$$

where \underline{I} is an identity matrix. The stiffness matrix, \underline{K} , becomes

$$\underline{K} = \int_V (\underline{d} \underline{f})^T \beta_2^2 \underline{D} (\underline{d} \underline{f}) dV \quad (37)$$

which is similar to the stiffness matrix of three-dimensional linear elasticity¹⁰. In the calculation of the change of natural frequency in a resonator due to a change in temperature, only the term $\beta_2^2 \underline{D}$ is updated which leads to an updating of the stiffness matrix.

The one-dimensional counterpart of the plate equation (34) represents either a strip or a beam. If $\partial(\cdot)/\partial x_3$ is set to zero, an equation for a strip of unit width is obtained:

$$\beta_2^2 \bar{D}_{11} u_{2,1111}^{(0)} = -\frac{3\rho}{b^2} \ddot{u}_2^{(0)} \quad (38)$$

If the moment in the width direction is set to zero ($t_{33}^{(1)} = 0$) and the corresponding strain $e_{33}^{(1)}$ is allowed to develop freely, eqn. (38) becomes a beam equation,

$$\beta_2^2 \bar{D}_{11} u_{2,1111}^{(0)} = -\frac{3\rho}{b^2} \ddot{u}_2^{(0)} \quad (39)$$

where $\bar{D}_{11} = \frac{S_{55}}{S_{11}S_{55} - S_{15}^2}$ is derived using the

same procedure used in arriving at eqns. (26). Solutions for eqns. (38) and (39) are compared with the finite element solutions for the Y-cut and NT-cut bars. It is interesting to note that the frequency-temperature behavior of eqns. (38) and (39) are functions of β_2^2 and

\bar{D}_{11} (or \bar{D}_{11}) only, since the density and dimensions are referred to a fixed reference frame. While the beam equation gives accurate predictions of the resonant frequencies, it is not adequate for studying frequency-

temperature characteristics.

FINITE ELEMENT ANALYSIS OF THE FREQUENCY-TEMPERATURE BEHAVIOR OF QUARTZ BARS.

An eight node hexahedron element with three degrees of translational freedom at each node was used. In order to "soften" an otherwise too "stiff" element, extra incompatible shape functions¹¹ were included in the element. A finite element package ANSYS was used in the analysis¹⁴. 2x2x2 Gauss quadrature integration points were used. 3x3x3 integration points were also employed but was found to yield the same results as the 2x2x2 integration points. Due to the large number of degrees of freedom in the eigenvalue problem of eqn. (12), the stiffness and mass matrices were reduced by choosing master degrees of freedom and statically condensing out the degrees of freedom not chosen using Guyan¹² reduction. There is no loss of accuracy in the reduced stiffness matrix, the approximation arises only in the reduced mass matrix¹². The accuracy in the frequency calculations depends upon whether the choice of the master degrees of freedom describes the salient displacement characteristics of the mode of vibration. The frequency change in parts per million, ppm, was calculated with reference to the frequency at 25°C ($= \frac{f - f_{25^\circ\text{C}}}{f_{25^\circ\text{C}}} \times 10^6$),

Fig. 1 shows the fundamental flexural mode of a clamped-free-clamped-free Y-cut plate with the length to thickness and width to thickness ratio, respectively, equal to 33.45 and 7.69. The dashed lines represents the undisplaced shape. A subspace iteration method¹⁵ was used to calculate the eigenfrequencies and mode shapes. The frequency-temperature curve was calculated and shown in fig. 2. The numerical curve is compared with the experimental curves by Paros and Singer and the strip, beam (eqns. (38) and (39)) and 3-branch curves. The 3-branch curve is the solution of a set of coupled one-dimensional plate equations calculated by Yong and Lee². The comparison between the numerical, Paros and 3-branch curves are good.

The first symmetrical flexural mode of an NT-cut beam is shown in fig. 3. Master degrees of freedom were chosen along the longitudinal axis through the center of the beam. We observe that in addition to the transverse flexure, the beam twists about the longitudinal axis. This coupling of the flexural vibrations with torsion may be the reason that the classical Bernoulli beam equation does not yield good predictions for the frequency-temperature behavior of NT-cut beams. This is shown in fig. 4 where the beam equation was used with three sets of temperature constants by Lee and Yong⁶,

Bechmann, Ballato and Lukaszek¹³, and Aruga¹. The finite element curve was calculated using the present constants by Lee and Yong⁶. Also shown is an experimental curve by Nakazawa, Nakamura and Miyashita¹. In fig. 5, the finite element result is compared with an empirical curve by Nakazawa, et al¹, a beam curve computed by them using Aruga's constants and a beam curve using eqn. (39) with the present temperature constants⁶. The finite element curves compare better with the experimental curves than the beam curves in figs. 4 and 5.

The quartz tuning forks are of particular interest because of their use in the watch industry. Fig. 6 shows a finite element model of a tuning fork. At the base, the nodes were constrained in all directions. The dashed lines in the figure show the undisplaced tines. Also given in the figure is the crystal orientation of the tuning fork. The turnover temperature, TZTC, changes with the angle of cut ψ about the X (electrical) axis. This change in the TZTC with the angle ψ is plotted in fig. 7 with the experimental curves by Nakazawa, et al¹ for a width to thickness ratio of 1.00. There are, in addition, two beam curves calculated by Nakazawa¹ et al employing temperature coefficients by Bechmann, Ballato and Lukaszek and Aruga. It must be noted that the beam equation cannot adequately describe the frequency-temperature characteristics of a tuning fork due to its complex boundary conditions and the coupling of flexure with torsion. If symmetrical vibrations about the crotch of the tuning fork were assumed, so that only one tine is modelled, the TZTC curve was found to shift upwards by 6°C. Furthermore, the length and shape of the base of the tuning fork has significant effects on its frequency-temperature characteristics.

REFERENCES

- 1) Nakazawa, M., Nakamura, Y and Miyashita, S., "Frequency-Temperature Characteristics of Quartz Crystal Flexure Bars and Quartz Crystal Tuning Forks," IEEE Trans. on Sonics and Ultrasonics, Vol. SU-26, No. 5, Sept. 1979, pp 369-376.
- 2) Yong, Y. K. and Lee, P. C. Y., "Frequency-Temperature Behavior of Flexural and Thickness-Shear Vibrations of Rectangular Rotated Y-cut Quartz Plates," Proc. of the 39th Ann. Symposium on Frequency Control, 1985, pp 415-426.
- 3) Tomikawa, Y., Sato, K., Konno, M. and Johnson, R. A., "Second-Mode Tuning Forks for High Frequencies. Finite Element Analysis and Experiments," IEEE Trans. on Sonics and Ultrasonics, Vol. SU-27, No. 5, Sept. 1980, pp 253-257.

- 4) Momosaki, E., Kogure, S., Inoue, M. and Sonoda, T., "New Quartz Tuning Fork with Very Low Temperature Coefficient," Proc. of the 33rd Ann. Symposium on Frequency Control, 1979, pp 247-254.
- 5) Kanbayashi, S., Okano, S., Hiram, K. and Kudama, T., "Analysis of Tuning Fork Crystal Units and Application into Electronic Wrist Watches," Proc. of the 30th Ann. Symposium on Frequency Control, 1976, pp 167-174.
- 6) Lee, P. C. Y and Yong, Y. K., "Frequency-Temperature Behavior of Thickness Vibrations of Doubly-Rotated Quartz Plates Affected by Plate Dimensions and Orientations," Proc. of the 38th Ann Symposium on Frequency Control, 1984, pp 164-175.
- 7) Mindlin, R. D., "An Introduction to the Mathematical Theory of Vibrations of Elastic Plates," U. S. Army Signal Corps Engineering Labs., Fort Monmouth, N. J, 1955.
- 8) Lee, P. C. Y., Wang, Y. S. and Harkenscoff, X., "High Frequency Vibrations of Crystal Plates Under Initial Stresses," J. Acoust. Soc. Am., Vol. 57, No. 1, Jan 1975, pp 95-105.
- 9) A. E. H. Love, A Treatise on the Mathematical Theory of Elasticity, (Dover, New York, 1944), pp 166.
- 10) W. Weaver and P. R. Johnston, Finite Elements for Structural Analysis, (Prentice-Hall, Englewood Cliffs, 1984), pp 10.
- 11) Wilson, E. L., Taylor, R. L., Doherty, W. P. and Ghaboussi, J., "Incompatible Displacement Models," Numerical and Computer Methods in Structural Mechanics, edited by S. J. Fenves, et. al., (Academic Press, New York, 1973) pp 42-57.
- 12) Guyan, R. J., "Reduction of Stiffness and Mass Matrices," AIAA Journal, Vol. 3, No. 2, Feb. 1965, pp 380.
- 13) Bechmann, R., Ballato, A. D. and Lukaszek, T. J., "Higher-Order Temperature Coefficients of the Elastic Stiffnesses and Compliances of Alpha-Quartz," Proc. IRE 50, 1962, pp 1812-1822.
- 14) ANSYS, Rev. 4.2, Swanson Analysis Systems Inc., Houston, Pennsylvania.
- 15) Bathe, K. J., Finite Element Procedures in Engineering Analysis, (Prentice-Hall, Englewood Cliffs, N. J. 1982), pp 672.

APPENDIX

Values of the temperature derivatives $D_{Pq}^{(n)}$ in $\text{N/m}^2/(\text{°C})^n$
 $Y\text{-Cut } (\theta=0, \phi=0)$

$D_{Pq}^{(1)}$:						
-0.592169D+07	-0.185382D+08	-0.839828D+07	-0.150867D+07	0.0	0.0	
		-0.839828D+07	0.150867D+07	0.0	0.0	
		-0.211767D+08	0.0	0.0	0.0	
			-0.107799D+08	0.0	0.0	
				-0.107799D+08	-0.150867D+07	0.630625D+07
$D_{Pq}^{(2)}$:						
-0.101190D+05	-0.221667D+05	-0.121760D+05	0.974577D+03	0.0	0.0	
	-0.101190D+05	-0.121760D+05	-0.974577D+03	0.0	0.0	
		-0.169184D+05	0.0	0.0	0.0	
			-0.157770D+05	0.0	0.0	
				-0.157770D+05	0.974577D+03	0.602384D+04
$D_{Pq}^{(3)}$:						
-0.537020D+01	0.137137D+02	-0.516465D+00	0.146149D+02	0.0	0.0	
	-0.537020D+01	-0.516465D+00	-0.146149D+02	0.0	0.0	
		0.100440D+02	0.0	0.0	0.0	
			-0.264388D+01	0.0	0.0	
				-0.264388D+01	0.146149D+02	-0.954195D+01

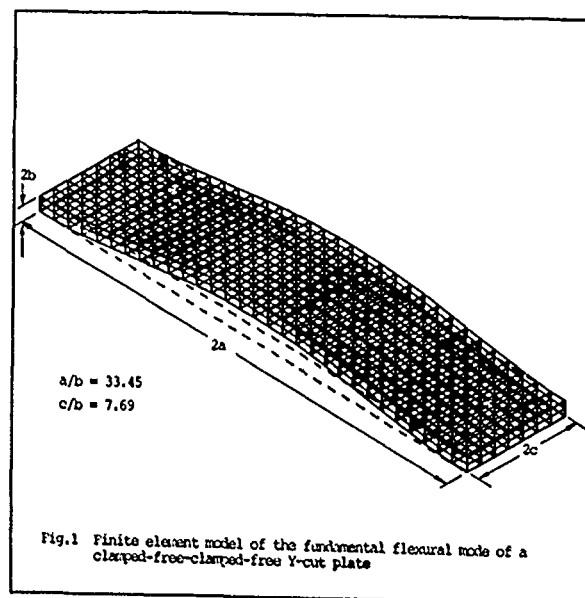


Fig.1 Finite element model of the fundamental flexural mode of a clamped-free-clamped-free Y-cut plate

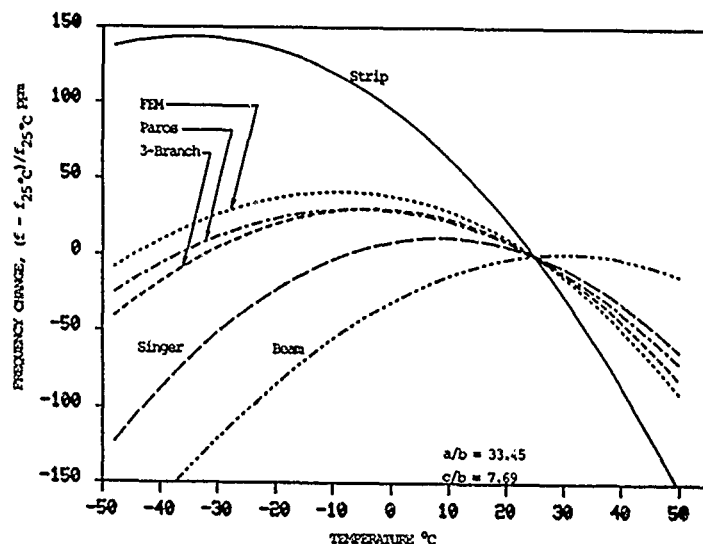
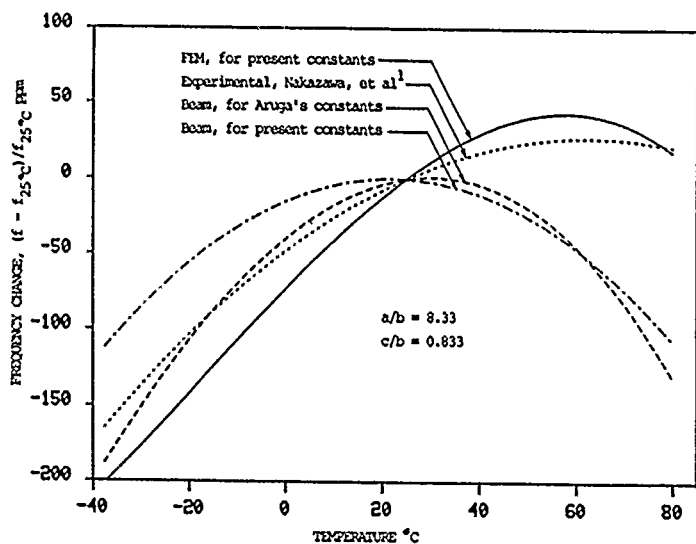
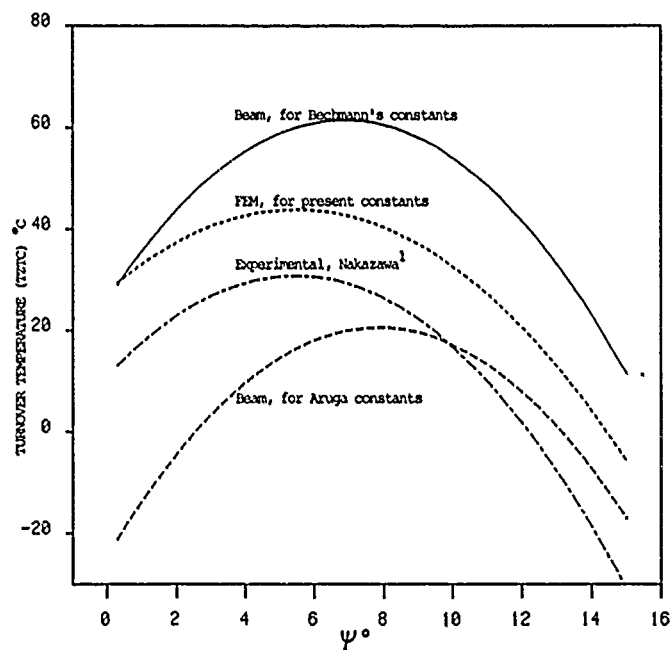
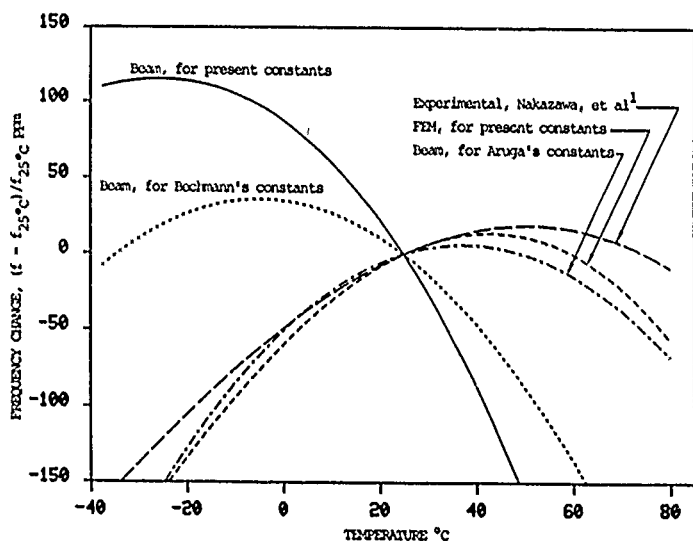
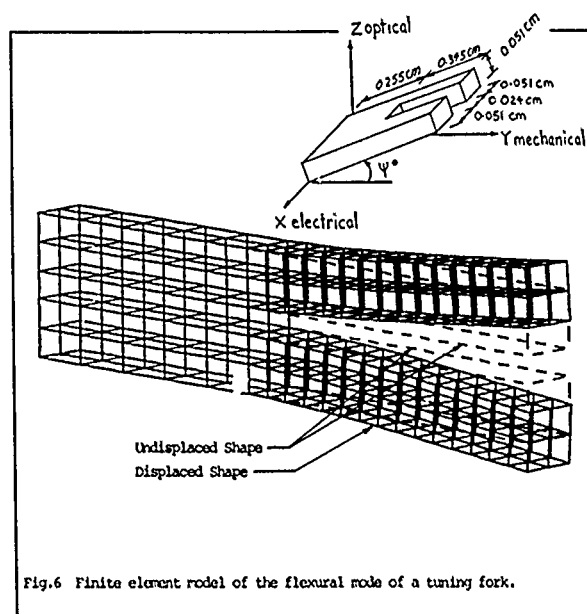
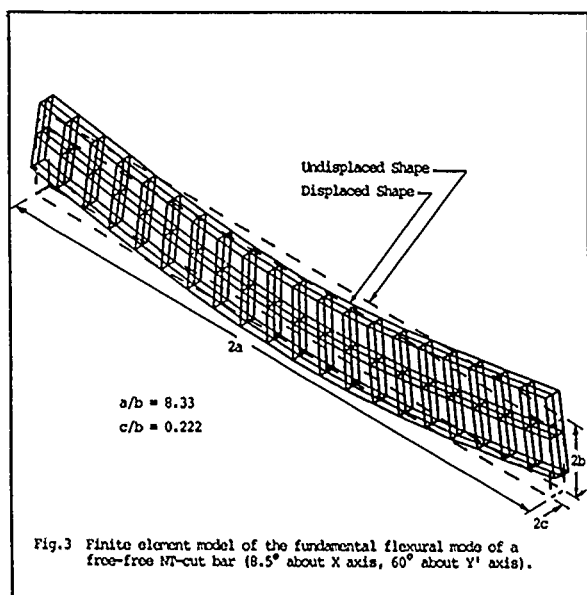


Fig.2 Frequency-temperature behavior of the Y-cut plate of fig.1.



**SUPPRESSION OF ANHARMONIC SPURIOUS MODES
BY MODIFIED ELECTRODE DESIGN USING CHARGE CANCELLATION**

Hitoshi SEKIMOTO and Toshiyuki IHARA
Faculty of Technology, Tokyo Metropolitan University, Tokyo, 158 Japan
Hozumi NAKATA and Masaaki MIURA
Nihon Dempa Kogyo Co., Ltd., Tokyo, 151 Japan

Abstract

The suppression of anharmonic spurious modes in trapped energy resonators with circular electrodes is theoretically studied. The suppression is accomplished by a modified electrode design using the effect of charge cancellation. A configuration of electrodes presented here consists of two portions; a central circular portion, and an outer ring portion with a larger value of mass loading. The motional inductances of third overtone AT-cut quartz resonators are calculated. The results indicate that the ring electrodes affect the vibrational distribution, and thus it is possible to charge cancel strong spurious modes by proper choices of the ring electrode mass loading and width. The electrode dimensions necessary to attain the inductance ratio over 1000 of spurious modes to the main mode are determined. It is also shown that this design may improve the motional inductances of the main modes below one-half of those obtained by the usual design.

I. Introduction

In AT-cut quartz resonators, strong spurious responses belong to the anharmonic thickness-shear family. Their suppression may be usually accomplished by using a small ratio of electrode diameter-to-plate thickness. However, crystal filter designers need frequently to choose large values of electrode diameter, in order to reduce the impedance level of a crystal filter. The undesirably large electroded area then causes a strong spurious to occur near in frequency to the main mode, and the design of crystal filters to be very complicated. Such a problem is common in the design of a filter employing third or fifth overtone crystals.

In this paper, the suppression of strong anharmonic modes in third overtone trapped energy resonators with circular electrodes is theoretically studied. The suppression is achieved by a modified electrode design using the effect of charge cancellation.

The effect of charge cancellation has been used for the selective suppression of anharmonic modes by R.J.Byrne et al. [1] and G.T.Pearman [2]. Byrne et al. showed the plate and electrode dimensions necessary to charge cancel a particular spurious mode in a rectangular plate. Also, Pearman investigated that the suppression of spurious modes in a rectangular beveled plate by the charge cancellation method is practical. But it is difficult to apply their techniques to overtone crystals. Because the vibrational deformation associated with strong anharmonic modes in overtone resonators is confined to the electroded area and so is not affected by the configuration of plates. In this paper, another approach is adopted. The vibrational deformation is strongly perturbed by a modified configuration of electrodes, that consists of two portions; a central circular portion, and a ring portion with a larger value of mass loading. The ring portion forces the position of displacement nodes in spurious modes and thus the total magnitude of

charges induced on the electrode to change. The shape of thickness-shear displacement, which determines the relative strength of any particular resonance, can be determined for infinite plates with such circular electrodes by extending an approximate theory given by H.Sekimoto [3].

The frequency, mode shape and motional inductances for third overtone resonators with the modified circular electrodes are calculated. The circular geometry of electrodes generates two spurious resonances near the main mode. They are close in frequency and furthermore the one, designated (3,1,3) anharmonic mode, is much weaker than the other, designated (3,3,1) anharmonic mode. This is due to small anisotropy between the X and Z' axes in the third overtone thickness-shear mode. Since the small anisotropy means that the wavelength along the plate is almost the same in the X and Z' directions, the ring portion of electrodes with a larger value of mass loading can effectively change the vibrational deformation of these trapped anharmonic modes. The calculated results of motional inductances indicate that the charge cancellation of each spurious mode occurs at two values of ring electrode width for a fixed total electrode diameter. It is possible, by proper choice of the ring electrode mass and width, to sufficiently suppress both the strong (3,3,1) and weak (3,1,3) anharmonic modes by electric charge cancellation. The electrode dimensions necessary to obtain the ratio more than 1000 of the spurious mode inductance to the main mode inductance are determined from the approximate theory. It is also shown that this design may improve the motional inductances of the main modes below one-half of those obtained by the usual design.

II. Approximate Equations

A schematic diagram of a trapped energy resonator with modified circular electrodes (density: ρ' , total radius: r_0) on an infinite AT-cut quartz plate (density: ρ , thickness: $2h$) is shown in Fig.1 where the x_1 , x_2 and x_3 coordinates are chosen as the X, Y' and Z' axes, respectively. The configuration of electrodes consists of two portions; a central circular portion (radius: r_c , thickness: $2h_c'$), and a ring portion (width: r_s , thickness: $2h_s'$) with a larger value of mass loading ($h_s' > h_c'$). This resonator has three separate cutoff frequencies for the ring electroded, central electroded, and surrounding unelectroded regions, designated ω_0^s , ω_0^c and ω_0' , respectively. The lowest cutoff frequency ω_0^s , as well as the ring electrode width, plays an important role in changing the vibrational field in the electroded region and thus producing the charge cancellation of spurious modes. The shape of thickness-shear displacements, which determines the relative strength of any particular resonance, can be determined by extending the previous work [3] on resonators with usual circular electrodes to resonators proposed here with modified circular electrodes.

According to Ref.[3], it is assumed that the x_2 dependence of displacement u_1 in each region is

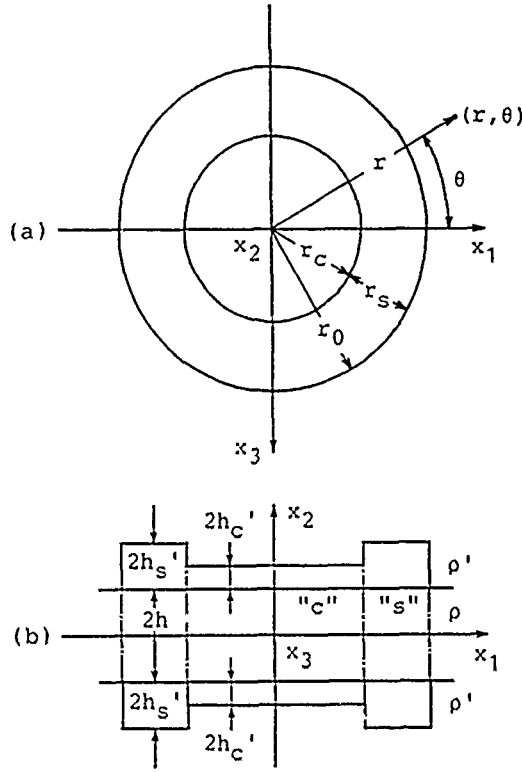


Fig.1 A trapped energy resonator with a modified configuration of circular electrodes.
(a) Top view. (b) Side view.

obtained from the pure thickness-shear solution at the cutoff frequency, and two-dimensional displacement field in the x_1x_3 plane near cutoff is expressed by the superposition of one-dimensional guided waves. That is

$$u_1 = U(r, \theta) \sin q_0 x_2 \quad (1)$$

where q_0 is the wavenumber in the x_2 direction at a cutoff ω_0 . $U(r, \theta)$ represents the amplitude at (r, θ) which denotes a polar position in the x_1x_3 plane, and is given by

$$U(r, \theta) = \int_C A_\psi e^{jk_\psi r} \cos(\psi - \theta) d\psi \quad (2)$$

$$\omega^2/\omega_0^2 - 1 = (\alpha_1 \cos^2 \psi + \alpha_3 \sin^2 \psi) k_\psi^2 h^2 \quad (3)$$

where A_ψ and k_ψ are the amplitude and wavenumber of the one-dimensional guided thickness-shear waves propagating at an angle ψ to the x_1 axis, respectively, and C is a path of integration [3]. Eq.(3) represents an approximate dispersion relation between k_ψ and frequency ω . Here α_i ($i=1$ or 3) is the curvature of a dispersion curve at ω_0 for thickness-shear waves propagating along the x_i axis, and can be obtained by using a numerical method [4] which takes into account the influence of anisotropy, piezoelectric stiffening, electrode mass loading and electrical short-circuiting.

Since the structure of resonator is symmetric with respect to the plane center, only the quadrant of the plane, i.e. $0 \leq \theta \leq \pi/2$, is taken into account. Furthermore, since we are interested in trapped energy modes piezoelectrically coupling a driving voltage across the electrodes, only the even modes in both x_1 and x_3 will be part of the analysis. Accordingly, from eqs.(2) and (3), the two-dimensional amplitude U in each region may be written with the Bessel functions of even order in the following form.

For the central electroded region ($0 < r < r_c$),

$$U^C(r, \theta) = \sum_{n=0}^{\infty} a_n^C J_{2n} \left(\frac{r}{h} k^C(\theta) \right) \cos 2n\theta_c$$

$$k^C(\theta) = \sqrt{\left(\frac{\omega^2}{\omega_0^2} - 1 \right) \left(\frac{\cos^2 \theta}{\alpha_1^C} + \frac{\sin^2 \theta}{\alpha_3^C} \right)}$$

$$\theta_c = \tan^{-1} (\sqrt{\alpha_1^C / \alpha_3^C} \tan \theta) \quad (4)$$

where a_n^C is the arbitrary constant, ω_0^C is the cutoff frequency, α_i^C ($i=1$ or 3) is the curvature of dispersion curve at ω_0^C , and J_{2n} denotes the first Bessel function.

For the ring electroded region ($r_c < r < r_0$),

$$U^S(r, \theta) = \sum_{n=0}^{\infty} \{ a_n^S J_{2n} \left(\frac{r}{h} k^S(\theta) \right) + b_n^S N_{2n} \left(\frac{r}{h} k^S(\theta) \right) \} \cos 2n\theta_s$$

$$k^S(\theta) = \sqrt{\left(\frac{\omega^2}{\omega_0^S} - 1 \right) \left(\frac{\cos^2 \theta}{\alpha_1^S} + \frac{\sin^2 \theta}{\alpha_3^S} \right)}$$

$$\theta_s = \tan^{-1} (\sqrt{\alpha_1^S / \alpha_3^S} \tan \theta) \quad (5)$$

where a_n^S , b_n^S are the arbitrary constants, ω_0^S is the cutoff frequency, and α_i^S ($i=1$ or 3) is the curvature of dispersion curve at ω_0^S , and N_{2n} denotes the second Bessel function.

For the unelectroded region ($r > r_0$),

$$U'(r, \theta) = \sum_{n=0}^{\infty} a_n' K_{2n} \left(\frac{r}{h} k'(\theta) \right) \cos 2n\theta'$$

$$k'(\theta) = \sqrt{\left(1 - \frac{\omega^2}{\omega_0'^2} \right) \left(\frac{\cos^2 \theta}{\alpha_1'} + \frac{\sin^2 \theta}{\alpha_3'} \right)}$$

$$\theta' = \tan^{-1} (\sqrt{\alpha_1' / \alpha_3'} \tan \theta) \quad (6)$$

where a_n' is the arbitrary constant, ω_0' is the cutoff frequency, and α_i' ($i=1$ or 3) is the curvature of dispersion curve at ω_0' , and K_{2n} denotes the second modified Bessel function.

The analysis of resonators requires mechanical boundary conditions for the present approximate solutions at discontinuities between adjacent regions. Since the true boundary conditions are that the displacements and normal components of stress are continuous at the discontinuity, it is reasonable to require that the similar conditions be constructed from the two-dimensional amplitude U . Ref.[3] shows that a two-dimensional force vector F associated with U approximately represents stress fields in the x_1x_3 plane. Thus, the boundary conditions are approximated by the continuity of

$$U, F_r \quad (7)$$

at junctions $r=r_c$ and r_0 . The normal component F_r of the force vector F is related to U by [3]

$$F_r(r, \theta) = \omega_0^2 h^2 H_0 \left\{ \alpha_1 \cos \theta \frac{\partial}{\partial x_1} U(r, \theta) + \alpha_3 \sin \theta \frac{\partial}{\partial x_3} U(r, \theta) \right\}$$

$$H_0 = \int_{\text{plate}} \rho \sin^2 q_0 x_2 dx_2 \quad (8)$$

where the integration in H_0 is carried out over the whole thickness of the plate including electrodes. Eq.(8) yields an expression with the Bessel functions of odd order in each region.

In computation of trapped energy fields and resonant frequencies, the a_{2n}^C , a_{2n}^S , b_{2n}^S and a_{2n}^I series of unknown constants must be truncated after some finite number of terms. Consequently, the field matching cannot be satisfied for any θ . We attempt to match the boundary conditions of eq.(7) only for some discrete points θ_n ($0 \leq \theta_n \leq \pi/2$) at $r=r_C$ and r_0 . If N terms are kept in each series and $2N$ matching points are selected, $4N$ linear homogeneous equations in $4N$ unknowns will be obtained. It is then possible to calculate approximate fields and resonant frequencies for trapped energy resonances.

When the vibrational field and the surface charge density on the electrode are known, the equivalent motional inductance that characterizes the relative strength of each resonant mode can be evaluated [5]. It is reasonable to assume in accordance with the approximation in eq.(1) that the charge distribution $q(r, \theta)$ on the electrode is proportional to U^C and U^S , that is

$$q = Q_0^C U^C, \text{ on the central electrode} \quad (9a)$$

where Q_0^C represents the free surface charge density on the central electrode associated with the pure thickness-shear vibration $u_1 = \sin q_0^C x_2$ at ω_0^C , and

$$q = Q_0^S U^S, \text{ on the ring electrode} \quad (9b)$$

where Q_0^S represents the free surface charge density on the ring electrode associated with the pure thickness-shear vibration $u_1 = \sin q_0^S x_2$ at ω_0^S . Accordingly, the equivalent inductance L is given by

$$L = \int_V \rho u_1^2 dV / \left[\int_S q ds \right]^2 \quad (10)$$

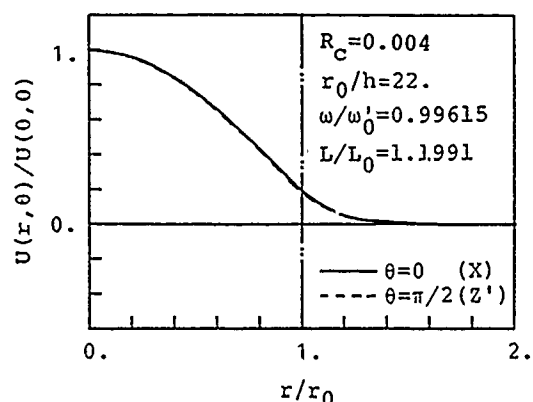
where V is the whole volume including the electrodes, and S is the whole area of the electrode.

This expression may be used to compute the motional inductances as a function of the central and ring electrode dimensions including the mass of electrodes with the vibrational fields determined at each resonance. $L \rightarrow \infty$ occurs when the denominator of eq.(10) goes to zero, that is, the positive and negative charges induced on the electrode exactly cancel. In usual resonators on infinite plates, the charge cancellation would never be observed because of the displacements necessarily decaying in the unelectroded region. In modified resonators presented here, however, it is expected that the ring portion of electrodes affect strongly the displacements under the electrodes and thus the charge cancellation of anharmonic spurious modes takes place.

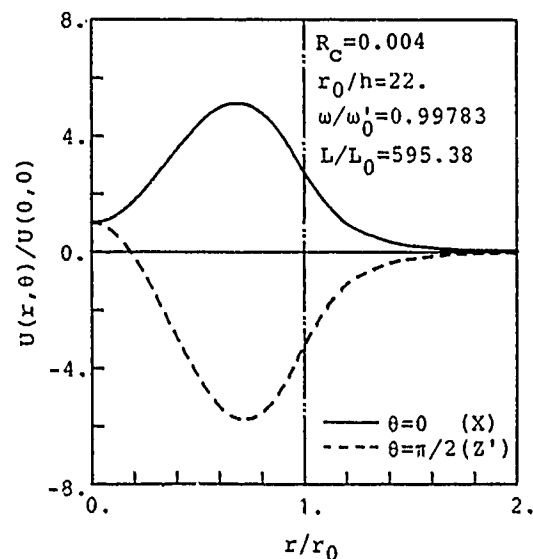
III. Suppression of spurious modes

In this section we treat the trapped energy resonators on an infinite AT-cut quartz plate operating in the third overtone mode. Before proceeding with the analysis of resonators with modified circular electrodes, the characteristics of anharmonic spurious modes on usual resonators without the ring electrodes are examined.

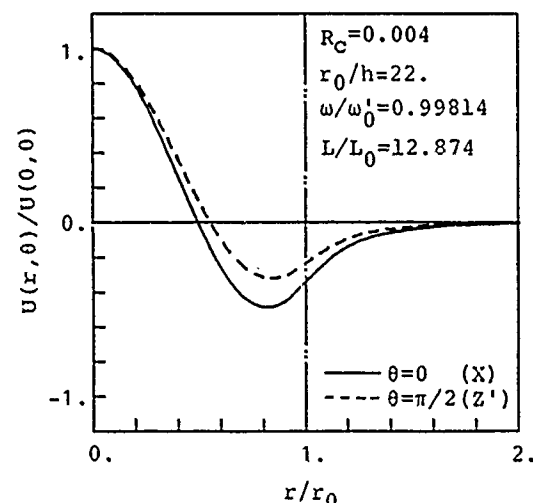
The calculated vibrational patterns $U(r, \theta)$ are shown in Fig.2 for the first three anharmonic modes, designated the main (3,1,1) mode and two spurious



(a) (3,1,1) mode



(b) (3,1,3) mode



(c) (3,3,1) mode

Fig.2 Calculated vibrational patterns in a third overtone AT-cut quartz resonator with circular electrodes for $r_C/h=r_0/h=22$, $r_S/h=0$.

(3,1,3) and (3,3,1) modes, where $r_C/h=r_0/h=22$, $r_S/h=0$ and the plateback $R_C (=2\rho'h_C'/\rho h)$ by the electrode plating equal to 0.004. This electrode diameter-to-plate thickness ratio $r_0/h=22$ for $R_C=0.004$ is about the maximum value permissible for only the two spurious modes, as well as the main mode, to exist. In

References

- [1] B.J.Byrne, P.Lloyd and W.J.Spencer, "Thickness-shear vibration in rectangular AT-cut quartz plates with partial electrodes," J. Acoust. Soc. Am., vol.43, No.2, pp.232-238, 1967
- [2] G.T.Pearman, "Thickness-twist vibration in beveled AT-cut quartz plates," J. Acoust. Soc. Am., vol.45, No.4, pp.928-934, 1969
- [3] H.Sekimoto, "Analysis of trapped energy resonators with circular electrodes," IEEE Trans. Sonics Ultrason., vol.SU-31, pp.664-669, 1984
- [4] H.Sekimoto and M.Ariga, "Dispersion characteristics near cutoff frequencies for thickness waves in high coupling piezoelectric plates," Trans. IECE Japan, vol.61-A, pp.980-987, 1978 (in Japanese)
- [5] R.Holland and E.P.EerNisse, Design of Resonant Piezoelectric Devices. Cambridge, Mass:MIT, 1969, p.136

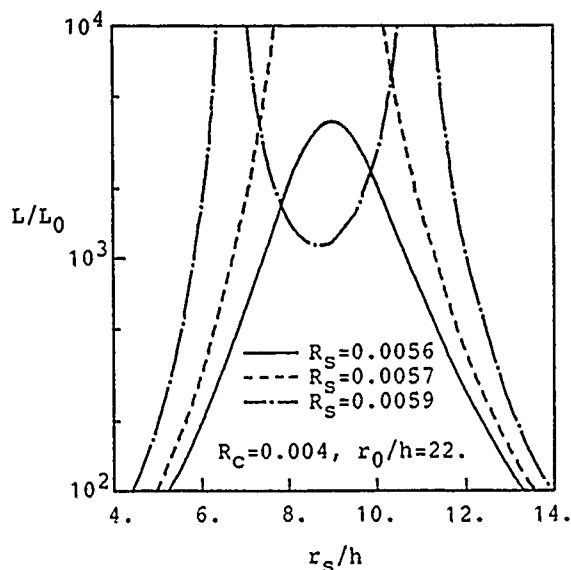


Fig.6 Calculated motional inductances of the (3,3,1) mode for various values of the ring electrode plateback R_S .

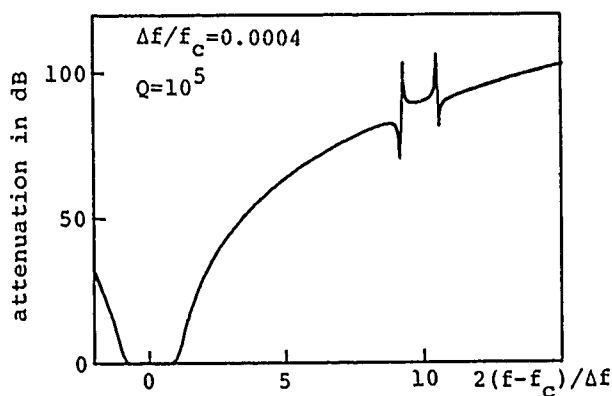


Fig.7 Calculated spurious response for a 4th order filter in form of cascaded semilattices with 3 dB relative bandwidth $\Delta f/f_c = 4 \times 10^{-4}$ and crystal $Q = 10^5$, where f_c , Δf represent the center frequency and 3 dB bandwidth, respectively. Two pairs of modified resonators with $r_0/h = 22$ are used.

will not be required in the design of filters. Therefore, it is reasonable to evaluate a relation between r_s/h and R_S necessary to suppress the (3,3,1) mode within a level below the main response. Since the spurious level depends on the bandwidth and structure of a filter to be designed, we took a fourth order filter in form of cascaded semilattices with 3 dB relative bandwidth 4×10^{-4} and crystal Q value 10^5 as a typical example, and numerically designed it using two pairs of modified resonators with $r_0/h = 22$. The theoretical filter shape is shown in Fig.7. Two spurious responses are suppressed below 60 dB. They are due to the (3,3,1) mode and their inductance ratios to the main (3,1,1) mode are 1600 and 3800. On the basis of similar computations, we estimated that a spurious level below 55 dB could be easily acquired for inductance ratios over 1000. Fig.8 shows a range of r_s/h and R_S necessary to obtain these ratios. The optimum region to give the maximum adjustable range of R_S will depend on the tolerance of r_s/h . The part enclosed with a dashed line represents an optimum region when a variation $\Delta(r_s/h) = \pm 1$ is assumed.

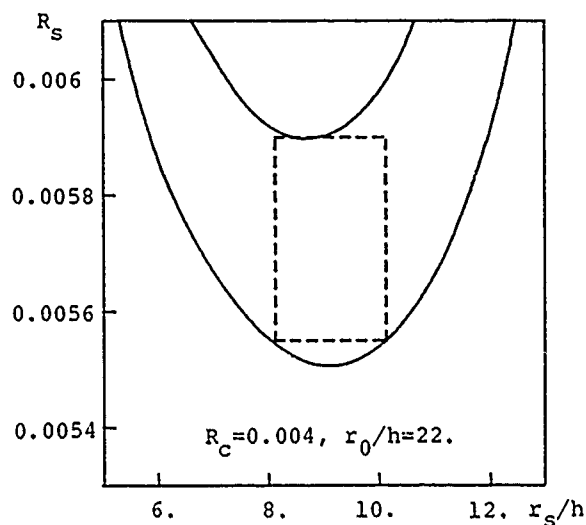


Fig.8 Diagrammatic representation of the ring electrode width r_s/h and plateback R_S for keeping the inductance ratio of the (3,3,1) mode to the main (3,1,1) mode more than 1000 for $r_s/h = 22$. A part enclosed a dashed line shows a region giving the maximum adjustable range of R_S for a deviation ± 1 of r_s/h .

Table I Optimum region giving the maximum adjustable range of R_S for a deviation ± 1 of r_s/h .

R_C	r_0/h	r_s/h	R_S	r_0'/h
.002	31.1	12.8 ± 1	.00280 ~ .00297	19.4
.003	25.4	10.4 ± 1	.00417 ~ .00444	16.2
.004	22.0	9.1 ± 1	.00555 ~ .00590	14.2
.005	19.7	8.1 ± 1	.00693 ~ .00736	12.8

r_0'/h : Maximum electrode diameter-to-plate thickness ratio for a resonator with a usual configuration of circular electrodes.

Finally, such optimum regions for the plateback $R_C = 0.002 \sim 0.005$ of the central electrodes were determined, provided that each value of r_0/h was set to the maximum value permissible for only the first two spurious (3,1,3) and (3,3,1) modes as well as the main (3,1,1) mode to exist. The results are listed in Table I, with the maximum electrode size r_0'/h for the usual resonator without the ring electrodes. A comparison between r_0/h and r_0'/h indicates that the present design using the effect of charge cancellation may improve the motional inductances of the main modes below one-half of those obtained by the usual design.

References

- [1] B.J.Byrne, P.Lloyd and W.J.Spencer, "Thickness-shear vibration in rectangular AT-cut quartz plates with partial electrodes," J. Acoust. Soc. Am., vol.43, No.2, pp.232-238, 1967
- [2] G.T.Pearman, "Thickness-twist vibration in beveled AT-cut quartz plates," J. Acoust. Soc. Am., vol.45, No.4, pp.928-934, 1969
- [3] H.Sekimoto, "Analysis of trapped energy resonators with circular electrodes," IEEE Trans. Sonics Ultrason., vol.SU-31, pp.664-669, 1984
- [4] H.Sekimoto and M.Ariga, "Dispersion characteristics near cutoff frequencies for thickness waves in high coupling piezoelectric plates," Trans. IECE Japan, vol.61-A, pp.980-987, 1978 (in Japanese)
- [5] R.Holland and E.P.EerNisse, Design of Resonant Piezoelectric Devices. Cambridge, Mass:MIT, 1969, p.136

VARIATIONAL ANALYSIS OF GT CUT QUARTZ CRYSTAL RESONATORS WITH THE SUPPORTING PORTIONS AT THE ENDS

Hirofumi KAWASHIMA

SEIKO ELECTRONIC COMPONENTS
1110 Hirai-cho Tochigi-shi Tochigi-ken, Japan 328

Summary

It is well known that a GT cut quartz crystal resonator coupled between two extensional mode vibrations has a cubic curve in frequency temperature characteristics as well as an AT cut quartz crystal resonator. In this paper, a study on frequency temperature characteristics and electrical characteristics of the GT cut quartz crystal resonator with the supporting portions at both ends of the vibrational portion has been made.

First, a secular equation is derived from a vibration analysis of the resonator consisting of the vibrational portion and the supporting portions, by applying the energy method. From this secular equation, a relationship of resonant frequency, frequency temperature coefficients versus a cut angle, a dimensional ratio and a mass ratio for the GT cut quartz crystal resonator is theoretically made clear. At the same time, as the result of comparing the calculated values with the experimental ones, it is confirmed that they agree well. Next, electrical characteristics of the quartz crystal resonators manufactured for trial experiments are measured, as a result, it is confirmed that a GT cut quartz crystal resonator with excellent frequency temperature characteristics, low crystal impedance and a high quality factor is realized.

Introduction

Recently, high precision, miniaturization, low consumption power and lightness are energetically being performed in the fields of consumer products and communication equipment. In the circumstances, quartz crystal resonators which are used in them have also the same trend. A GT cut quartz crystal resonator is a resonator satisfying these requisites, which is comparatively of low frequency and most excellent in frequency temperature characteristics as a single quartz crystal resonator. However, it is not for practical use except a part, because of difficulty of miniaturization and support for the conventional rectangular GT cut quartz crystal resonator.

In order that the authors overcome the above-mentioned drawbacks, we have proposed the new shape GT resonator¹. This resonator makes it possible to overcome the drawbacks of the conventional GT plate. In designing this resonator, the finite element method is available for analyzing vibration modes of the resonators, and the new shape GT cut quartz

crystal resonator with low energy loss, a high quality factor and excellent frequency temperature characteristics is obtained in the experiments. However, a theoretical analysis on the frequency temperature characteristics has by no means been performed.

In order that the author makes clear on frequency temperature characteristics of a GT cut quartz crystal resonator consisting of the vibrational portion and the supporting portions theoretically, first, a secular equation is calculated from partial differential equations derived from an energy method with Galerkin's method. From this frequency equation obtained, a relationship of resonant frequency, frequency temperature coefficients versus a cut angle, a dimensional ratio and a mass ratio is made clear. Next, the calculated values are compared with the experimental ones. Finally, electrical characteristics of GT cut quartz crystal resonators manufactured for trial experiments are shown.

Derivatives of the frequency equation

Equation of motion

Fig.1 shows a coordinate system for a GT plate. This plate is shown in IRE expression of $yzwt(\phi/\theta)$, and the GT plate has the cut angle of $\phi=51.5^\circ$ and $\theta=45^\circ$. Fig.2 shows an actual resonator shape(a) consisting of a vibrating portion and supporting portions and its modified shape(b). It is thinkable that the resonator(a), as shown in Fig.2(b), has the supporting portions vibrating in a flexural mode with each mass m at both ends of the vibrational portion and also with the boundary conditions of "both hinged ends". That is to say, the supporting portions have each mass m and resistance to a vibration of vibrational portion (work done by external force). Let us take width x_0 and length z_0 in a dimension of the vibrational portion for this modified shape(b), which correspond to rotated crystal axes x' and z' of quartz respectively, furthermore, length L and width W in a dimension of the supporting portions and the center of the vibrational portion to be the original point O .

Now, when taking kinetic energy K_1, K_2 of the vibrational portion and one supporting portion, potential energy U_1 of the vibrational portion, and work W_2 done by external force of the supporting portions, the total kinetic energy K and the total potential energy U are as follows.

$$K = K_1 + K_2 I_{+z_0/2} + K_2 I_{-z_0/2} \quad (1)$$

$$U=U_1 \quad (2)$$

Therefore, Lagrangian L is obtained.

$$\mathcal{L}=K-U \quad (3)$$

In addition, since the resistance caused by the supporting portions can be considered as external force operated on the vibrational portion, the work W done by the external force is expressed as follows.

$$\hat{W}=W'_2|_{+z_0/2}+W'_2|_{-z_0/2} \quad (4)$$

Therefore, when taking time t, from the variational principle²,

$$\delta \int_{t_0}^{t_1} \mathcal{L} dt + \int_{t_0}^{t_1} \delta \hat{W} dt = 0 \quad (5)$$

When thickness y_0 of the resonator is very thin as compared with another dimension ($x_0 \gg y_0$, $z_0 \gg y_0$), the kinetic energy K_1 of the vibrational portion is expressed by the following equation(6).

$$K_1 = \int_{-x_0/2}^{x_0/2} \int_0^{y_0} \int_{-z_0/2}^{z_0/2} \frac{1}{2} \rho \left\{ \left(\frac{\partial u}{\partial t} \right)^2 + \left(\frac{\partial w}{\partial t} \right)^2 \right\} dx dy dz \quad (6)$$

where, ρ is density of quartz, u and w are the displacement in the x and z axis direction respectively. The variation of the equation (6) integrated in time is performed and by part integration,

$$\delta \int_{t_0}^{t_1} K_1 dt = - \int_{t_0}^{t_1} dt \left[\int_{V_1} \rho \left(\frac{\partial^2 u}{\partial t^2} \delta u + \frac{\partial^2 w}{\partial t^2} \delta w \right) dV \right] \quad (7)$$

$$\text{where, } \int_{-x_0/2}^{x_0/2} \int_0^{y_0} \int_{-z_0/2}^{z_0/2} dx dy dz = \int_{V_1} dV$$

Furthermore, the kinetic energy K_2 of one supporting portion is given as follows.

$$K_2 = \frac{1}{2} m \left(\frac{\partial w'}{\partial t} \right)^2 \quad (8)$$

where, m shows a mass of the supporting portion and w' is the displacement of the supporting portion. As well as the equation(6), the variation of the equation(8) is as follows, by part integration.

$$\delta \int_{t_0}^{t_1} K_2 \Big|_{\pm z_0/2} dt = -m \int_{t_0}^{t_1} dt \left[\frac{\partial^2 w'}{\partial t^2} \delta w' \right]_{\pm z_0/2} \quad (9)$$

Next, let us consider the potential energy U_1 of the vibrational portion. A relationship between stress T_i and strain S_i ($i=1,3,5$) is given as follows³.

$$\left. \begin{aligned} T_1 &= r_{11} S_1 + r_{13} S_3 + r_{15} S_5 \\ T_3 &= r_{31} S_1 + r_{33} S_3 + r_{35} S_5 \\ T_5 &= r_{51} S_1 + r_{53} S_3 + r_{55} S_5 \end{aligned} \right\} \quad (10)$$

where, if taking elastic compliance s_{ij} ($i,j=$

$$\left[\begin{array}{ccc} r_{11} & r_{13} & r_{15} \\ r_{31} & r_{33} & r_{35} \\ r_{51} & r_{53} & r_{55} \end{array} \right] = \left[\begin{array}{ccc} s_{11} & s_{13} & s_{15} \\ s_{31} & s_{33} & s_{35} \\ s_{51} & s_{53} & s_{55} \end{array} \right]^{-1} \quad (11)$$

1,3,5) of the plate, r_{ij} is given as a function of the s_{ij} .

Therefore, the potential energy U_1 is calculated from the following equation.

$$U_1 = \int_{V_1} \frac{1}{2} (T_1 S_1 + T_3 S_3 + T_5 S_5) dV \quad (12)$$

where, if taking the integration term of the equation(12) to be U_0 ,

$$U_0 = \frac{1}{2} (T_1 S_1 + T_3 S_3 + T_5 S_5) \quad (13)$$

then, since the U_0 is a function of strain S_1 , S_3 and S_5 , δU_0 is as follows.

$$\delta U_0 = \frac{\partial U_0}{\partial S_1} \delta S_1 + \frac{\partial U_0}{\partial S_3} \delta S_3 + \frac{\partial U_0}{\partial S_5} \delta S_5 \quad (14)$$

In addition, if taking a small area in the arbitrary direction of the plate to be dS and its perpendicular direction to be n ,

$$\left. \begin{aligned} dy dz &= dS \cos(x, n) \\ dx dy &= dS \cos(z, n) \end{aligned} \right\} \quad (15)$$

Accordingly, the variation of the equation (12) is given as follows, by using the equations(10), (14) and (15).

$$\begin{aligned} \delta \int_{t_0}^{t_1} U_1 dt &= \int_{t_0}^{t_1} dt \left[\int_S \left[(T_1 \cos(x, n) \right. \right. \\ &\quad \left. \left. + T_3 \cos(z, n) \right) \delta u \right. \\ &\quad \left. + (T_5 \cos(x, n) + T_3 \cos(z, n)) \delta w \right] dS \\ &\quad - \int_{V_1} \left[\left(\frac{\partial T_1}{\partial x} + \frac{\partial T_3}{\partial z} \right) \delta u \right. \\ &\quad \left. + \left(\frac{\partial T_3}{\partial x} + \frac{\partial T_5}{\partial z} \right) \delta w \right] dV \end{aligned} \quad (16)$$

Next, the work W_2 done by external force of the supporting portion is given by the following relationship⁵.

$$W'_2 = \int_{-L/2}^{L/2} \frac{1}{2} EI \left(\frac{\partial^2 w'}{\partial x^2} \right)^2 dx \quad (17)$$

where, E is young's modulus, I is moment of inertia. Therefore, the variation of the equation(17) is calculated with the same method mentioned above and obtained as follows.

$$\begin{aligned} \int_{t_0}^{t_1} \delta W'_2 \Big|_{\pm z_0/2} dt &= \int_{t_0}^{t_1} dt \left[EI \left(\frac{\partial^2 w'}{\partial x^2} \right) \delta \left(\frac{\partial w'}{\partial x} \right) \Big|_{-L/2}^{L/2} \right. \\ &\quad \left. - EI \left(\frac{\partial^3 w'}{\partial x^3} \right) \delta w' \Big|_{-L/2}^{L/2} \right. \\ &\quad \left. + \int_{-L/2}^{L/2} EI \left(\frac{\partial^4 w'}{\partial x^4} \right) \delta w' \Big|_{\pm z_0/2} dx \right] \end{aligned} \quad (18)$$

Furthermore, since the supporting portion can be considered with the boundary conditions of "both hinged ends" at both ends, from the equation(18) $z_0 \gg \lambda \gg 0$, See Fig.3),

$$\begin{aligned} \frac{\partial^2 w'}{\partial x^2} &= 0 & ; x = \pm L/2, z = \pm z_0/2 \\ w' &= 0 \end{aligned}$$

Therefore,

$$w' = \pm w_0 \cos \frac{\pi}{L} x \sin \eta \frac{z_0}{2} \cos \omega' t \Big|_{z=\pm z_0/2} \quad (19)$$

where, w_0 is amplitude, η is a constant which is obtained from the boundary conditions, ω' is angular frequency. In addition, the external force $F_{\pm z_0/2}$ is calculated from the equation (18) using the equation (19).

$$\begin{aligned} F'_{\pm z_0/2} &= - \int_{-L/2}^{L/2} EI \left(\frac{\partial^4 w'}{\partial x^4} \right) dx \\ &= \mp w_0 EI \left(\frac{\pi}{L} \right)^3 \sin \eta \frac{z_0}{2} \cos \omega' t \end{aligned} \quad (20)$$

The force given in the equation (20) is the force at $z = \pm z_0/2$ and $|x| \leq \delta_1 \neq 0$ (δ_1 is width in the x axis direction at the connected portion of the vibrational portion and the supporting portion), and it operates on the vibrational portion in the z axis direction as resistance.

Therefore, a distribution of force applied to the vibrational portion is given by $F_{\pm z_0/2} \cos \eta x$ (η is a constant which is determined by the boundary conditions). Furthermore, since the force operating at $\pm z_0/2$ is the same, and this force can be considered as the force per unit volume \bar{F}_{\pm} operating on the vibrational portion, accordingly, the following relationship is obtained.

$$\bar{F}_{\pm} = \frac{2F_{\pm z_0/2} \cos \eta x}{V_1} \quad (21)$$

Substituting the equations (7), (9), (16) and (18) into the equation (5), since δu , δw and $\delta w'$ are respectively arbitrary under the condition that the vibrational portion and the supporting portions are connected at $z = z_0/2$ and $|x| \leq \delta_1 \approx 0$, the equations of motion are given as follows. In addition, the boundary conditions are quite

$$\left. \begin{aligned} \rho \frac{\partial^2 u}{\partial t^2} &= \frac{\partial T_1}{\partial x} + \frac{\partial T_2}{\partial z} \\ \rho \frac{\partial^2 w}{\partial t^2} \pm \bar{F}_{\pm} &= \frac{\partial T_3}{\partial x} + \frac{\partial T_4}{\partial z} \end{aligned} \right\} \quad (22)$$

similarly obtained from the equation (5).

$$\begin{aligned} T_1 = T_2 = 0 & ; x = \pm x_0/2 \\ T_3 = 0 & ; z = \pm z_0/2 \\ \int_{\delta_1} T_3 dS &= \mp m \frac{\partial^2 w'}{\partial t^2} \Big|_{\pm z_0/2} ; z = \pm z_0/2, |x| \leq \delta_1 \neq 0 \end{aligned} \quad (23)$$

Galerkin's method

Now for, let's try to solve the equations of the motions for the equation (22). The displacement u and w can be expressed as follows, including a term of time.

$$\left. \begin{aligned} u &= U'(x, z) \cos \omega t \\ w &= W'(x, z) \cos \omega t \end{aligned} \right\} \quad (24)$$

Furthermore, though a distribution of displacement for the vibrational portion connected to the supporting portions added is much complicated, by means of the finite element method, namely, from the analysis with the finite element method, the following relation is obtained by adding the approximate calculation between the displacement and the coordinate axes x, z .

$$\left. \begin{aligned} U'(x, z) &= A_1 \sin \xi x \cosh \eta' z \\ W'(x, z) &= -A_3 \cos \xi x \sin \eta z \end{aligned} \right\} \quad (25)$$

where η' is a constant. In addition, since the vibrational portion and the supporting portions are connected at $z = \pm z_0/2$ and $|x| \leq \delta_1 \approx 0$, from the equations (24), (25) and (19),

$$w = w' ; z = \pm z_0/2, |x| \leq \delta_1 \neq 0$$

Therefore, $w_0 = -A_3$, $\omega = \omega'$. Furthermore, \bar{F}_{\pm} in terms of (21) is given as follows from the equation (20).

$$\bar{F}_{\pm} = \pm \bar{F}_0 \cos \omega t \quad (26)$$

where, \bar{F}_0 has the following relation.

$$\bar{F}_0 = \frac{2A_3}{V_1} EI \left(\frac{\pi}{L} \right)^3 \sin \eta \frac{z_0}{2} \cos \xi x$$

The resistance of the equation (26) shows elongation of the vibrational portion produced by force \bar{F}_{\pm} . If taking the elongation quantity \bar{w} , the \bar{w} can be expressed as follows:

$$\begin{aligned} \bar{w} &= \bar{W}(x) \cos \omega t \\ \bar{W}(x) &= \bar{F}_0 / r_{33} \eta^2 \end{aligned}$$

Therefore, the displacement w is given as follows, taking the resistance (very weak) into account.

$$w = (\bar{W} + W') \cos \omega t \quad (27)$$

From the equations (10), (24), (25), (27), and taking into consideration $S_5 = 0$ from piezoelectric nature for the GT quartz crystal plate, the equation (22) is transformed into the following form.

$$\left. \begin{aligned} -\rho \omega^2 U' &= r_{11} \frac{\partial^2 U'}{\partial x^2} + r_{13} \frac{\partial^2 W'}{\partial x \partial z} + r_{31} \frac{\partial^2 U'}{\partial z \partial x} + r_{33} \frac{\partial^2 W'}{\partial z^2} \\ -\rho \omega^2 (\bar{W} + W') &+ \bar{F}_0 = r_{11} \frac{\partial^2 U'}{\partial x^2} + r_{13} \frac{\partial^2 W'}{\partial x \partial z} + r_{31} \frac{\partial^2 U'}{\partial z \partial x} \\ &+ r_{33} \frac{\partial^2 W'}{\partial z^2} \end{aligned} \right\} \quad (28)$$

Next, taking the boundary conditions into account, since $S_5 = 0$ from the piezoelectric nature, \bar{F}_0 is calculated from the first term of the equation (23).

$$\xi = \frac{a\pi}{x_0} \quad (a = 1, 3, 5, \dots) \quad (29)$$

From the third term, the following result is

obtained.

$$\eta = \frac{2}{z_0} \zeta \quad (30)$$

where, ζ is a solution of $\cot \zeta = \mu \zeta$, if taking the sectional area A° perpendicular to the z axis, there is a relation of $\mu = m / (\rho A^\circ z_0 / 2)$. In more detail, μ shows a ratio of a mass m of a supporting portion to a half mass of a vibrational portion, namely, a mass ratio. Furthermore, since γ_{51} and γ_{53} are approximately zero for the GT quartz crystal plate from the second term, T_5 is also zero approximately. η' satisfies the boundary conditions of the equation(23) at an arbitrary value, however, here, analyzing displacement U' of a width-extensional vibration by the finite element method and from this analyzed result, an optimum value η' is calculated. As a result, if taking a constant ζ , η' is expressed as follows.

$$\eta' = \frac{2}{z_0} \zeta' \quad (31)$$

Accordingly, displacement U' , W' satisfying the boundary conditions from ζ , η and η' , is obtained.

Next, let's solve the equation(28). Since the equation(28) is not easily solved, an approximate solution by means of Galerkin's method is performed. That is to say, multiplying the first term of the equation(28) by the function U' satisfying the boundary conditions and the second term by $(W+W')$ as well, and by performing integration over area $(-x_0/2 \leq x \leq x_0/2, -z_0/2 \leq z \leq z_0/2)$. The calculated result gives the following simultaneous equations.

$$\left. \begin{aligned} (\rho \omega^2 - r_{11} \xi^2) A_1 + r_{13} \xi \eta \Phi_{13} A_3 &= 0 \\ r_{31} \xi \eta' \Phi_{31} A_1 + (\rho \omega^2 - r_{33} \eta'^2) A_3 &= 0 \end{aligned} \right\} \quad (32)$$

where, Φ_{13} and Φ_{31} are calculated as follows.

$$\Phi_{13} = \frac{4\zeta'}{\zeta^2 + \zeta'^2} \frac{(\zeta \sin \zeta \cosh \zeta' + \zeta' \cos \zeta \sinh \zeta')}{(2\zeta' + \sinh 2\zeta')} \quad (33)$$

quite similarly,

$$\Phi_{31} = \frac{-4\zeta}{\zeta^2 + \zeta'^2} \frac{(\zeta \cos \zeta \sinh \zeta' - \zeta' \sin \zeta \cosh \zeta')}{(2\zeta - \sinh 2\zeta') + F_{ND} \zeta} \quad (34)$$

where, F_{ND} is given as follows.

$$F_{ND} = \frac{\pi^3}{24} \frac{E}{r_{33}} \frac{z_0}{x_0} \left(\frac{W}{L} \right)^3 \frac{1}{\zeta^2} \sin \zeta \quad (35)$$

In addition, since $A_1 \neq 0$ and $A_3 \neq 0$ from the equation(32), the following relation is obtained.

$$\left| \begin{array}{cc} \rho \omega^2 - r_{11} \xi^2 & r_{13} \xi \eta \Phi_{13} \\ r_{31} \xi \eta' \Phi_{31} & \rho \omega^2 - r_{33} \eta'^2 \end{array} \right| = 0 \quad (36)$$

From the above-mentioned relation, the frequencies are given.

$$f^2 = \frac{1}{2} \{ (f_w^2 + f_l^2) \pm \sqrt{(f_w^2 - f_l^2)^2 + 4 f_{cp}^4} \} \quad (37)$$

where,

$$f_w^2 = \frac{1}{4\pi^2} \frac{r_{11}}{\rho} \xi^2, \quad f_l^2 = \frac{1}{4\pi^2} \frac{r_{33}}{\rho} \eta'^2$$

$$f_{cp}^4 = \frac{1}{4\pi^2} \frac{r_{13}^2}{\rho^2 x_0^2 z_0^2} \zeta \zeta' \Phi_{13} \Phi_{31} \quad (\because r_{13} = r_{31})$$

Calculation of frequency temperature coefficients

Resonant frequencies of the equation(37) are a function of dimensions x_0 , z_0 , density ρ and elastic stiffness, and also these are a function of temperature T . Now, performing Taylor's series at the temperature T_0 concerning $f(T)$, the $f(T)$ is given.

$$f(T) = f(T_0) \{ 1 + \alpha(T - T_0) + \beta(T - T_0)^2 + \gamma(T - T_0)^3 + \dots \} \quad (38)$$

where, temperature coefficients α , β , γ are defined as follows.

$$\left. \begin{aligned} \alpha &= \left[\frac{1}{f(T)} \frac{\partial f(T)}{\partial T} \right]_{T=T_0} \\ \beta &= \left[\frac{1}{2! f(T)} \frac{\partial^2 f(T)}{\partial T^2} \right]_{T=T_0} \\ \gamma &= \left[\frac{1}{3! f(T)} \frac{\partial^3 f(T)}{\partial T^3} \right]_{T=T_0} \end{aligned} \right\} \quad (39)$$

Thus, these temperature coefficients α , β , γ are calculated from the equation(39). In this theoretical calculation they are calculated at $T_0 = 20^\circ\text{C}$. In addition, the constants of the reference(4) are used in the calculation.

Comparison of the calculated and experimental values

The equation(37) gives two frequencies. Therefore, let us take frequencies f_+ , f_- to the plus and minus signs respectively. Fig.3 shows enlarged detail of the supporting portion for the resonator illustrated in Fig.2. It is needless to say that, by the change of the vibrational portion, and also that, by the change of dimensions of the supporting portions, resonant frequency, frequency temperature characteristics and so on also change. The various characteristics are concretely shown below.

Frequency spectrum

Fig.4 shows a relationship between resonant frequencies f_+ , f_- and a dimensional ratio $r = x_0/z_0$ of a quartz crystal resonator which performs fundamental vibrations for two extensional modes, which has the cut angle of $\theta = 50^\circ$, $\theta = 45^\circ$ and the dimension of $\delta_1 = 60 \mu\text{m}$, $A = 80 \mu\text{m}$, $C = 0.7 \text{ mm}$, $D = 0.92 \text{ mm}$, $B = 80 \mu\text{m}$, $x_2 = 1.41 \text{ mm}$, namely, the mass ratio of $\mu = 6.33 \times 10^{-2}$, and calculating η' from the displacement U' analyzed by means of the finite element method, as a result, which has $\zeta = 1.30$ obtained from the equation(31). The solid lines in Fig.4 are the calculated values obtained by this theoretical analysis, and the circles are the calculated values by means of the

finite element method. Both calculated results for f_+ of the fundamental vibration agree well, while for f_- of the sub-vibration they show somewhat higher frequency in the finite element method than that in this theoretical analysis, however, the tendency between them agrees very well. Fig.5 shows a relationship of the frequency constant K_- of the sub-vibration which is determined by the dimension z_0 versus a mass ratio μ for a quartz crystal resonator with the cut angles of $\theta=51^\circ$, $\theta=45^\circ$ and the dimensional ratio $r=0.945$. The solid line is the calculated value and the circles are the experimental values. As the mass ratio increases, the frequency constant K_- decreases, the calculated and experimental values agree well. Thus, the frequency f_- of the sub-vibration is determined by the mass ratio μ . Fig.6 shows a relationship between a cut angle θ and frequencies f_+ , f_- of a quartz crystal resonator with the dimensional ratio $r=0.945$, $\mu=4.99 \times 10^{-2}$, $\theta=45^\circ$. The solid lines are the calculated values and the circles are the experimental values. As the cut angle increases, the frequencies f_+ , f_- decrease. These results also agree very well in the calculation and the experiments.

Frequency temperature coefficients and cut angle

Fig.7 shows a relationship between a cut angle θ and frequency temperature coefficients α , β of a quartz crystal resonator with the dimensional ratio $r=0.96$, the mass ratio $\mu=6.33 \times 10^{-2}$, the thickness $y_0=50 \mu\text{m}$ and $\theta=45^\circ$. The solid line and the dashed line are the calculated values and the signs O and X show the experimental values. According to this theoretical calculation, two cut angles of $\theta=52.1^\circ$ and $\theta=55.5^\circ$ whereat $\alpha=0$ are given. This calculated result agrees very well with the experimental values. On the other hand, with respect to the second order temperature coefficient β , this calculation gives the cut angle of $\theta=51.8^\circ$ whereat $\beta=0$, as the cut angle θ gets large, β also gets large. Though it seems that the difference between the calculated value and the experimental values exists, because of taking the scale of 10^{-9} for β , so as to contrast well it with the α , they comparatively agree well.

Fig.8 shows a relationship between a cut angle and the first order temperature coefficient α of a quartz crystal resonator with $r=0.96$, $\theta=45^\circ$, when taking a mass ratio μ as a parameter. As is apparent in Fig.7, the temperature coefficient α versus a cut angle shows a convex curve. In addition, as the mass ratio μ gets small, the first order temperature coefficient α inversely gets large at the same cut angle. This is because a coupling between the frequencies f_w , f_l gets stronger, as the frequency f_l gets high, according to decrease of the mass ratio μ .

For example, this calculation does not give any cut angle whereat $\alpha=0$ at $\mu=8.75 \times 10^{-2}$ and always the negative values. Next, when having a smaller value of $\mu=6.33 \times 10^{-2}$, it gives two cut angles of $\theta=51.5^\circ$ and $\theta=55.5^\circ$ whereat $\alpha=0$, in addition, when having a much smaller value of $\mu=4.99 \times 10^{-2}$, it gives two cut angles of $\theta=49.3^\circ$ and $\theta=57.7^\circ$ whereat $\alpha=0$ as well. Thus, the cut angle θ where gives $\alpha=0$ greatly changes by the value of the mass ratio μ .

Fig.9 shows a relationship between a cut angle and the third order temperature coefficient γ of a quartz crystal resonator with $\theta=45^\circ$, $\mu=6.33 \times 10^{-2}$, when taking a dimensional ratio $r=x_0/z_0$ as a parameter. The solid lines are the calculated values and the circles are the experimental values at $r=0.96$. Causing the big difference between the calculated and experimental values, it is thinkable because the constants⁴ used in this calculation have an error of 100 percent. However, the tendency versus the cut angle between the calculated values and the experimental ones agrees well. In addition, as a dimensional ratio r increases, the temperature coefficient γ at the same cut angle θ also increases.

Frequency temperature coefficients and dimensional ratio

Fig.10 shows a relationship between a dimensional ratio and the temperature coefficients α , β of a quartz crystal resonator with $\theta=51^\circ$, $\theta=45^\circ$. The change of the dimensional ratio r also causes that of a mass ratio, when a dimension of the supporting portions is constant. Therefore, the calculation of the mass ratio μ at each dimensional ratio r has been made, that is to say, the mass ratio μ of $\mu=5.86 \times 10^{-2}$, 5.91×10^{-2} , 5.97×10^{-2} is obtained at $r=0.94$, 0.95 , 0.96 . Furthermore, according to increase of the dimensional ratio, α , β also increase, the calculated values and the experimental values for α , β agree well.

Fig.11 shows a relationship between the dimensional ratio r and the first order temperature coefficient α of a quartz crystal resonator with $\theta=52^\circ$, $\theta=45^\circ$, when taking the mass ratio μ as a parameter. As the dimensional ratio r increases, α also increases, the value of α is very dependent on the mass ratio μ . For example, at $\mu=8.75 \times 10^{-2}$, α always gets the negative value at the dimensional ratio between 0.94 and 0.98 , at $\mu=6.33 \times 10^{-2}$, the calculation gives $\alpha=0$ at $r=0.96$, in addition, at $\mu=4.99 \times 10^{-2}$, it gives $\alpha=0$ at $r=0.948$. Thus, the dimensional ratio r where gives $\alpha=0$ greatly changes by the value of the mass ratio μ .

Frequency temperature characteristics

Fig.12 shows frequency temperature characteristics of a quartz crystal resonator for which α and β almost reach zero. The circles are the experimental values and the solid line is the calculated value. According to the experiments, since the quartz crystal resonator has α and β which almost reach zero under the conditions of $\theta=51.5^\circ$, $\theta=45^\circ$, $r=0.940$ and $\mu=4.99 \times 10^{-2}$, it shows the excellent frequency temperature characteristics, while, in this calculation, it shows the excellent frequency temperature characteristics under the conditions of $\theta=51.5^\circ$, $\theta=45^\circ$, $r=0.950$ and $\mu=4.99 \times 10^{-2}$ (α , β have the values of $\alpha=7.85 \times 10^{-8}/^\circ\text{C}$, $\beta=1.5 \times 10^{-11}/^\circ\text{C}^2$ and $\gamma=-4.95 \times 10^{-11}/^\circ\text{C}^3$ respectively). The cut angles θ , θ agree extremely well, however, though causing about one percent difference as regards the dimensional ratio r , those results agree well. In addition, comparing and examining the calculated value and the experimental value, they agree well in the temperature range of -10°C to $+50^\circ\text{C}$, however, they have somewhat difference in the wider temperature range,

because of influence of the third order temperature coefficient, that is to say, because γ in the calculation has a larger value than that in the experiments. Thus, it has been confirmed that the excellent frequency temperature characteristics are obtained both in the calculation and in the experiments

Various electrical characteristics

Table 1 shows typical values of electrical equivalent circuit constants for a quartz crystal resonator which has a dimension of width $x_0=1.56$ mm, a dimensional ratio of $r=0.945$, thickness of $y_0=51$ μ m and a mass ratio of $\mu=4.99 \times 10^{-2}$, and the electrodes disposed on the entire surface of the vibrational portion so as to excite two extensional vibration modes. In the Table 1, the signs show motional inductance L_1 , motional capacitance C_1 , shunt capacitance C_0 and series resistance R_1 .

As is obvious from the Table 1, a quartz crystal resonator with low crystal impedance and high quality values can be obtained, even though added the supporting portions. At the same time, the quartz crystal resonator of this study is excellent even in various electrical characteristics as compared with an AT cut quartz crystal resonator of the same frequency.

Conclusion

In this paper, the various theoretical and experimental examination has been made concerning a GT cut quartz crystal resonator coupling between two extensional vibration modes, so as to apply to consumer products and communication equipment. First of all, equation of motion of a resonator one-body-formed with the vibrational portion and the supporting portions is derived from an energy method. From the equations, the frequency equation is calculated by applying Galerkin's method to solve partial differential equations. Secondly, analyzing resonant frequency and frequency temperature characteristics, a relationship between a cut angle, a dimensional ratio, a mass ratio and frequency temperature coefficients has been made clear. Thirdly, as a result of comparing and examining the calculated values and the experimental values of resonators manufactured for trial experiments, it has been confirmed that the resonant frequency and the frequency temperature coefficients obtained from this theoretical analysis agree very well with those in the experiments. Finally, it has been also confirmed that the various electrical characteristics of this resonator are available for consumer products and communication equipment.

Acknowledgements

I wish to express my thanks to Mr. Otski and Mr. Kudo for their help in the experiments and also to Mrs. Watanabe and Miss. Watarai for their beautiful drawings. I would like to thank Messrs. Kubota, Kamata, Ishii and Kawamura for their help and encouragement.

References

1. H. Kawashima, H. Sato and O. Ochiai: "New frequency temperature characteristics of miniaturized GT cut quartz resonator", Proceeding of the 34th Annual Symposium on Frequency Control, pp. 131-139(1980).
2. A.E.H. Love: "A treatise on the mathematical theory of elasticity", Dover-New York(1934).
3. M. Nakazawa and S. Kozima: "A study of GT type quartz crystal plate", IEEE Trans. Sonics & Ultrason., SU 29, 3, pp. 121-127 May(1982).
4. M. Aruga: "On the elastic constants of quartz and their temperature characteristics", Bull. Tokyo Inst. Tech., Vol. A-2, pp. 88-182, 1956(in Japanese).
5. K. Matsudaira: "Fundamental Vibration", Kyoritsu Publishing(1950), (in Japanese).

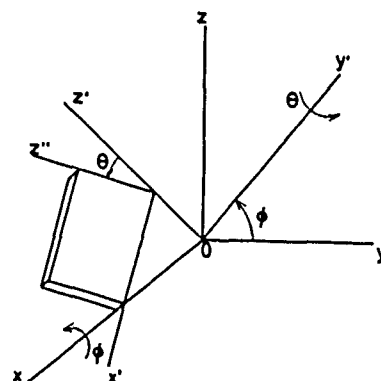


Fig.1 Coordinate system for GT plate.

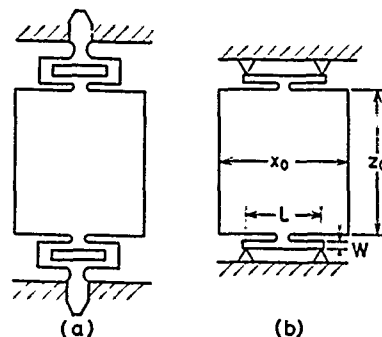


Fig.2 Resonator shape (a) and its modified shape (b).

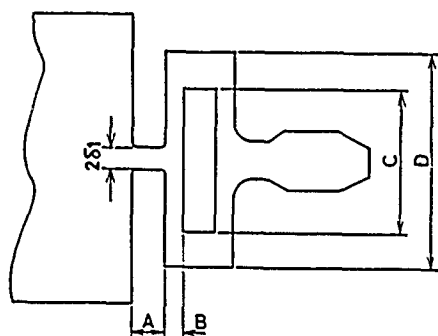


Fig.3 Enlarged detail of the supporting portion.

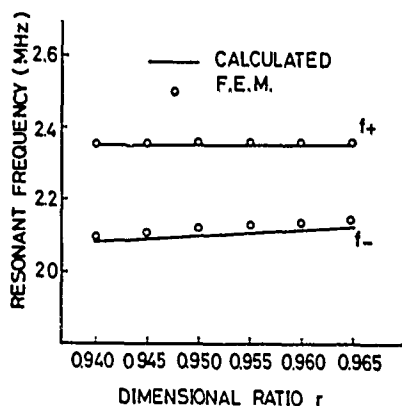


Fig.4 Frequency spectrum.

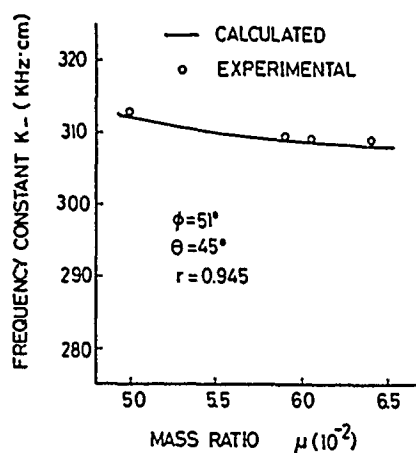


Fig.5 Relationship of mass ratio and frequency constant.

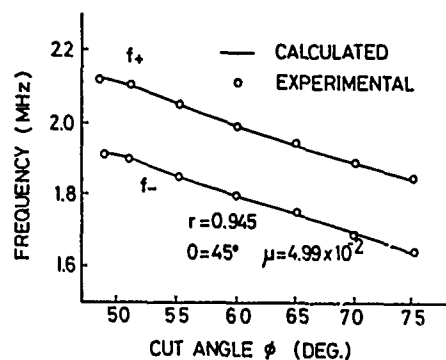


Fig.6 Relationship of cut angle and frequency.

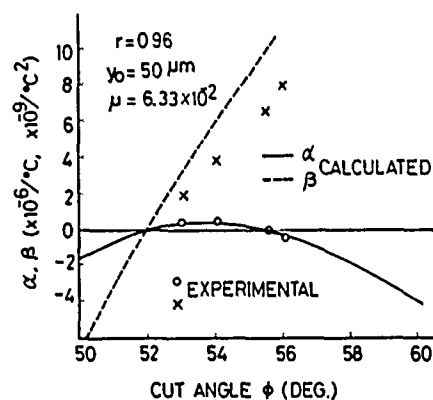


Fig.7 Relationship of cut angle and temperature coefficients α, β .

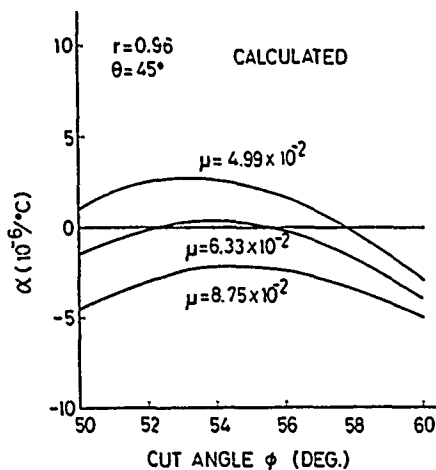


Fig.8 Relationship between cut angle and α when changing mass ratio.

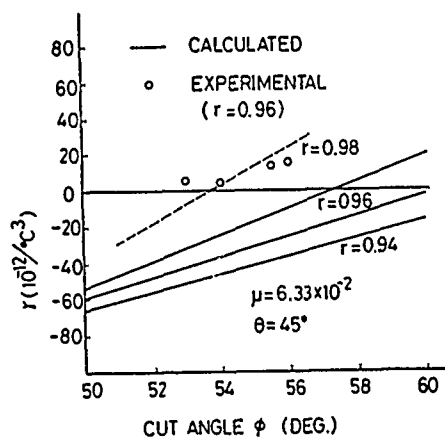


Fig.9 Relationship between cut angle and r .

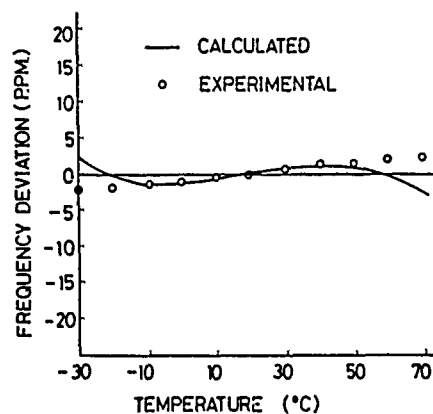


Fig.12 Frequency temperature characteristics.

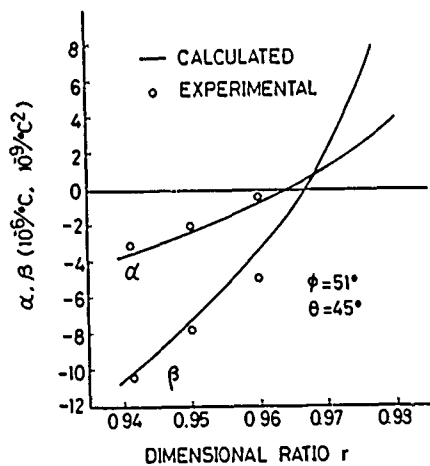


Fig.10 Relationship of dimensional ratio and temperature coefficients.

Table 1 Equivalent electrical circuit constants.

($\theta = 45^\circ$)						
Angle ϕ	Freq. (MHz)	R_1 (Ω)	L_1 (H)	C_1 (fF)	C_0 (pF)	Q ($\times 10^4$)
51°	2.10	75	0.97	5.90	2.40	17.0
55°	2.05	80	1.20	5.00	2.46	19.5

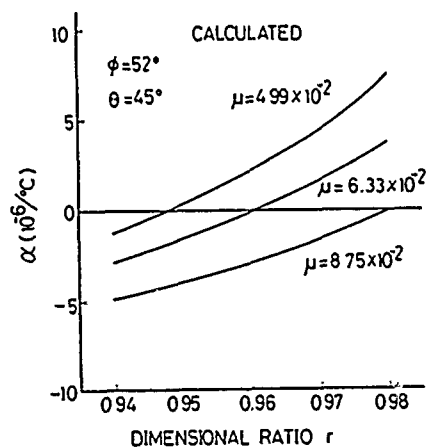


Fig.11 Relationship of dimensional ratio and α when changing mass ratio.

K. HIRAMA, T. SHOJI AND Y. TANAKA

Crystal Division, Toyo Communication Equipment Co., Ltd.
Samukawa-machi, Koza-gun, Kanagawa Pref., Japan

Abstract

This paper describes an AT-cut overtone mode crystal resonator that generates the desired overtone mode frequency in a coilless oscillator circuit.

Energy trapping theory shows that the energy trapping coefficient decreases as the overtone order decreases; consequently, the trapped energy of the lower overtone modes may be weaker than that of desired overtone mode. As poorly trapped energy dissipates in the mounting portion, the energy of lower order overtone modes is significantly more damped than that of the desired overtone mode.

A pair of main and subelectrodes is employed to realize the desired damping effect. The subelectrode is located either in the mounting portion or in the portion near to the edge. One key design parameter is the gap between the main electrode and the subelectrodes, which is determined so that the maximum ratio of the equivalent resistance at the lower order overtone modes to that at the desired overtone mode may be achieved.

Experimental results in 65MHz third-order overtone mode and 108MHz fifth-order overtone mode resonators, have shown that the equivalent resistance of lower overtone modes is at least several times larger than that of the desired overtone mode.

Introduction

AT-cut crystal resonators are the only resonators yet known in which both the first and second order coefficients of the frequency-temperature characteristic are close to zero at room temperature. Therefore, modern telecommunications owes a great debt to the appearance of these devices.

As telecommunication loads have grown, overtone-mode crystal resonators have been developed to permit communications in the higher frequencies of the VHF band. While it is impossible to produce fundamental-mode resonators with a frequency over several tens of megahertz, overtone-mode resonators have been able to fill this gap. To extract a particular overtone, however, requires a specially designed oscillation circuit with a negative impedance peak at the overtone frequency. Unfortunately, this means the use of an LC circuit, which is bulky and requires calibration. The use of LC circuits also presents barriers to the introduction of LSIs.

This paper introduces a new resonator design with a highly frequency-dependent crystal impedance that extracts the desired overtone frequency without the need for an LC circuit. The resonator is an AT-cut crystal with a minimum impedance falling right at the desired overtone frequency.

Analysis

Energy Trapping Characteristics of an Infinite Plate

Fig. 1 shows a resonator model used to analyze fundamental and overtone-mode energy trapping in the X direction of an infinitely long plate.

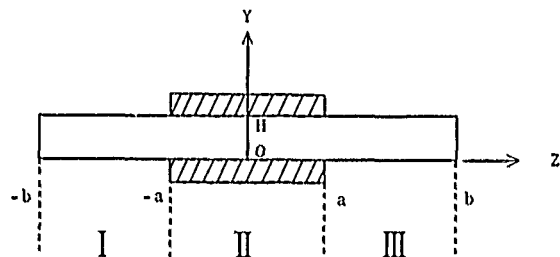


Fig. 1 A model for analyzing resonator behavior.

Restricting our analysis to shear horizontal (SH) waves parallel to the X axis, we arrive at the following expression for the displacement:

$$U = u \cos\left(\frac{n\pi y}{h}\right) \exp(j\omega t) \quad (1)$$

u is a function of z , as follows:

$$U_I = B \frac{\cosh k'z}{\sinh k'a} - C \frac{\sinh k'z}{\cosh k'a} \quad (-b \leq z \leq -a) \quad (2)$$

$$U_{II} = A \frac{\cos kz}{\sin ka} \quad (-a \leq z \leq a) \quad (3)$$

$$U_{III} = B \frac{\cosh k'z}{\sinh k'b} + C \frac{\sinh k'z}{\cosh k'b} \quad (a \leq z \leq b) \quad (4)$$

where the propagation constants k and k' are real in each of the three regions.

$$k = \frac{n\pi}{h} \sqrt{\left(\frac{f}{f_0}\right)^2 - 1} \quad (5)$$

$$k' = \frac{n\pi}{h} \sqrt{1 - \left(\frac{f}{f_0}\right)^2} \quad (6)$$

f_0 is the resonant frequency of a plate where region II is infinite, f_0' is the resonant frequency of a plate where regions I and III are infinite, and n is the order of the selected overtone. The upper part in this duplex recording corresponds to the symmetrical modes generated by the odd-order inharmonic overtones, while the lower part corresponds to the antisymmetrical modes of the even-order inharmonic overtones.

At $z = \pm a$, the displacement and stress are continuous; at $z = \pm b$, stress is free. These boundary conditions lead to the following formula for the frequency:

$$\tanh k'(b-a) = \frac{k}{k'} \tan ka \quad (7)$$

When f_0 , the resonant frequency of region II, is slightly lower than f_0' , that of regions I and III, we can approximate the frequency formula given in Eq. (7). First, let

$$f = (1 + \delta)f_0 \quad (8)$$

$$f' = (1 + \Delta)f_0 \quad (9)$$

Then we can write

$$\frac{k'}{k} = \sqrt{\frac{(1 - \psi)}{\psi}} \quad (10)$$

where $\psi = \frac{\delta}{\Delta}$ (11)

This gives us the following approximation for the frequency:

$$\begin{aligned} & \tanh \frac{n\pi a}{H} \cdot \sqrt{2\Delta(1 - \psi)} \cdot \left(\frac{b}{a} - 1\right) \\ &= \sqrt{\frac{1 - \psi}{\psi}} \tan \frac{n\pi a}{H} \sqrt{2\psi\Delta} \quad (12) \end{aligned}$$

Since b/a is a parameter, the frequency ψ is a function of $na/H \cdot \sqrt{\Delta}$.

Fig. 2 shows the resonant frequency of an energy-trapping resonator for the inharmonic modes from order s_0 to s_3 , where the parameter b/a is four. For comparison, the resonant frequency of an infinite plate is shown by a dotted line. The difference in response between finite and infinite plates is insignificant for inharmonic modes of all orders above the lowest inharmonic mode s_0 . There is some difference in the s_0 mode, but its qualitative behavior is similar.

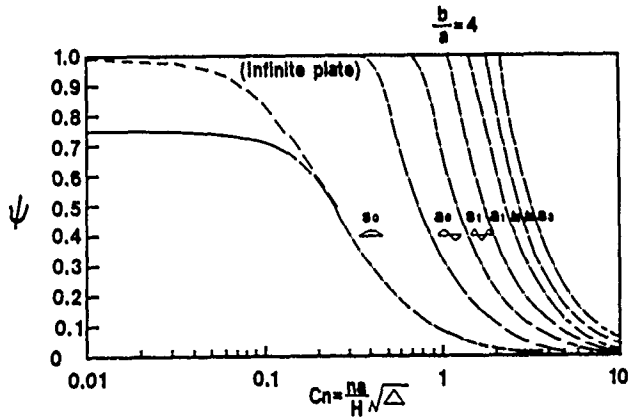


Fig. 2 Resonant frequencies of trapped modes in a finite plate.

Suppressing the Lower Harmonic Modes

Based on the preceding analysis, we determined the relationship between the normalized energy-trapping frequency of the plate ψ and the energy-trapping coefficient Cn . In a plate with a particular cutoff frequency, energy will be trapped in local regions where the cutoff frequency is lower. The amount of trapped energy, however, varies with the value of ψ : when $\psi = 1$, the trapped energy is zero; when $\psi = 0$, the energy is completely trapped. For ψ values between zero and one, the trapped energy varies continuously. Cn is given by the following equation:

$$Cn = \frac{na}{H} \sqrt{\Delta} \quad (13)$$

For any given resonator, the electrode dimension a , the plate thickness H , and the difference in cutoff frequency Δ are fixed. The overtone order n , on the other hand, is controlled by the overtone at which the

resonator is used. This means that Cn increases proportionally with the harmonic order n .

For small Cn values, the trapped energy rises with the overtone order; for large Cn values, the trapped energy is large for all overtone orders. Therefore, by judiciously selecting the value of Cn , we can achieve excellent trapping of the desired overtone order and above, which will minimize the crystal impedance at the frequency we want. The lower orders will not be trapped substantially, so they will tend to dissipate in the resonator mount.

The Effect of Subelectrodes and the Loss Angle

Our analysis suggested by adding a pair of subelectrodes we could effectively damp the vibration-energy leakage of the lower order overtones. Consider an infinite plate with the electrode pattern shown in Fig. 3.

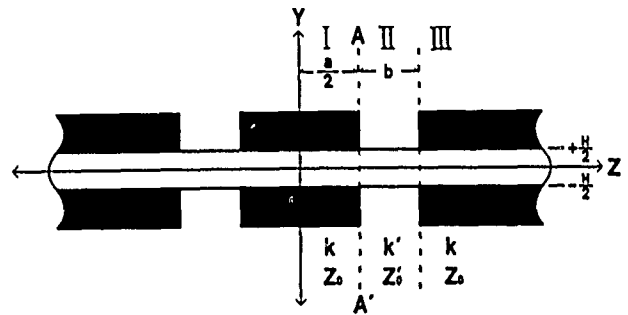


Fig. 3 An infinite plate with a subelectrode.

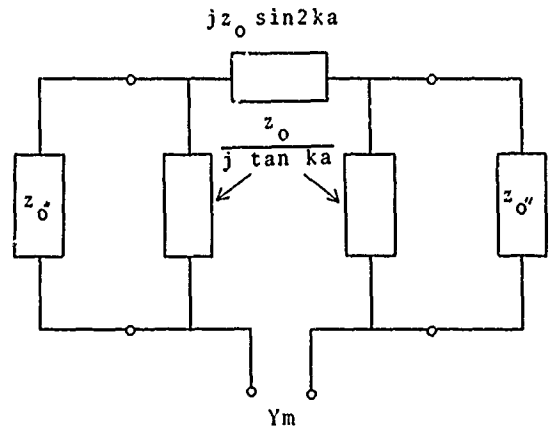


Fig. 4 Equivalent circuit.

Regions I and III have the same cutoff frequency, while that of region II is slightly higher. The electrode width is $a/2$, and the electrode spacing b . If we let the acoustic impedance from A to A' to the right side be $Z_{0''}$, we can describe it by the equivalent circuit of Fig. 4. The purely elastic admittance, could then be written as follows:

$$Y_m = \frac{\left(\frac{1}{Z_{0''}} + \frac{j}{Z_0} \tan ka\right) \frac{1}{jZ_0 \sin ka}}{\frac{1}{Z_{0''}} - \frac{j}{Z_0} \cot ka} \quad (14)$$

To see the effect of the subelectrodes, we calculate the acoustic impedance from A to A' .

$$z_0'' = \frac{z_0}{\cosh^2 k'b + \left(\frac{z_0}{z_0'}\right)^2 \sinh^2 k'b} - jz_0' \frac{\left(1 + \left(\frac{z_0}{z_0'}\right)^2\right) \tanh k'b}{1 + \left(\frac{z_0}{z_0'}\right)^2 \tanh^2 k'b} \quad (15)$$

The first term is the loss. Even if the resonator model is lossless, this term results from the vibration energy that propagates as a traveling wave through the infinitely large region III with cutoff frequency f_0 .

To determine the amount of loss, we define the "loss angle." In principle, this should be defined by the entire resonator, but for simplicity, we define it as the ratio of the real and imaginary parts of the acoustic impedance on the facet from A to A', which provides a straightforward means of qualitative analysis.

$$\tan \phi = \frac{z_0}{z_0'} \frac{2}{\left\{1 - \left(\frac{z_0}{z_0'}\right)^2\right\} \sinh 2k'b} \quad (16)$$

Next, let us determine the value of z_0/z_0' . From Eq. (10), we have

$$\sqrt{\frac{1 - \psi}{\psi}} = \frac{k'}{k} \quad (17)$$

Therefore,

$$\frac{z_0}{z_0'} = \sqrt{\frac{\psi}{1 - \psi}} \quad (18)$$

Substituting Eq. (18) into Eq. (16) gives us the loss angle.

$$\tan \phi = \frac{2}{\sinh 2k'b} \sqrt{\frac{\psi}{1 - \psi}} \quad (19)$$

Next, we express the k' part of Eq. (19) in terms of ψ . In principle, this should be done by deriving the resonant frequency from Fig. 4 and using this relationship between k' and ψ , but again, to simplify qualitative analysis, we used the approximation of an infinite plate with a single electrode. This approximation illustrates the major features of resonator behavior, and is reasonably accurate in cases where region II is very long. In this case, the frequency equation is as follows:

$$\tan ka = \frac{k'a}{ka} \quad (20)$$

Therefore,

$$k'b = \frac{b}{a} \sqrt{\frac{1 - \psi}{\psi}} \tan^{-1} \sqrt{\frac{1 - \psi}{\psi}} \quad (21)$$

$$\tan \phi = \underbrace{\frac{2}{\sinh \frac{b}{a} \sqrt{\frac{1 - \psi}{\psi}} \tan^{-1} \sqrt{\frac{1 - \psi}{\psi}}}}_I \cdot \underbrace{\sqrt{\psi(1 - \psi)}}_II \quad (22)$$

ϕ : Loss angle
 ψ : Normalized frequency

The first term is a monotonic function of the normalized frequency ψ . The second term has its maximum value in the interval $0 < \psi < 1$. This implies that there is a normalized frequency ψ that results in the maximum loss angle. This loss angle also increases proportionally with the ratio b/a of the main electrode width a to the electrode gap b , so a small gap b will increase the loss angle.

This makes the spacing between the main electrodes and subelectrodes a critical design factor. If the separation between the main and subelectrodes is too narrow, all the resonant energy, including that of the desired overtone mode, will leak to the mount and dissipate, resulting in a large equivalent resistance for all modes; if the spacing is too large, there will be no damping of any of the resonant modes. This suggests that between these two extremes we could find an optimum value which results in no loss of the desired overtone mode, while steeply attenuating all the harmonic modes below it.

Experimental Analysis

Based on the preceding qualitative analysis, we performed a number of experiments.

CI vs. Cn

Fig. 5 shows the relationship between the crystal impedance and Cn. The solid line is the resistance at the desired overtone. The dotted line is the resistance at the fundamental frequency, and indicates peak attenuation in the neighborhood of $Cn = 2.4$. The corresponding Cn value at the fundamental mode is $Cn = 0.8$ -- an excellent match with the qualitative analysis, which indicated this same value.

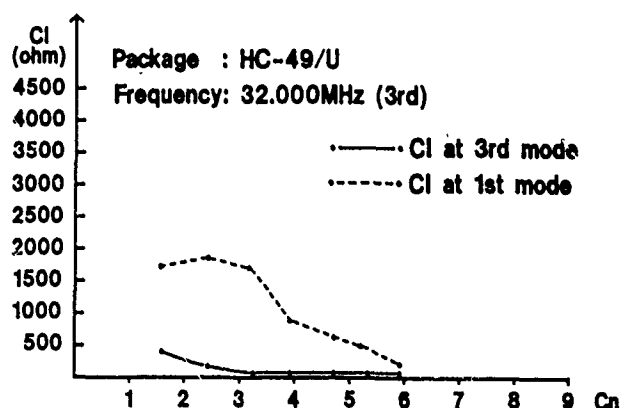


Fig. 5 The relationship between crystal impedance (CI) and the energy trapping coefficient (Cn).

CI vs. Gap

Fig. 6 shows the effect of the gap between the main and subelectrodes on the crystal impedance. The gap was cut successively by a YAG laser and then CI was measured. The gap, shown on the X-axis, is expressed as the ratio (%) of the main electrode width to the electrode gap. You can see that there is an optimum gap value where the attenuation of the lower harmonics is a maximum and attenuation of the desired overtone is a minimum.

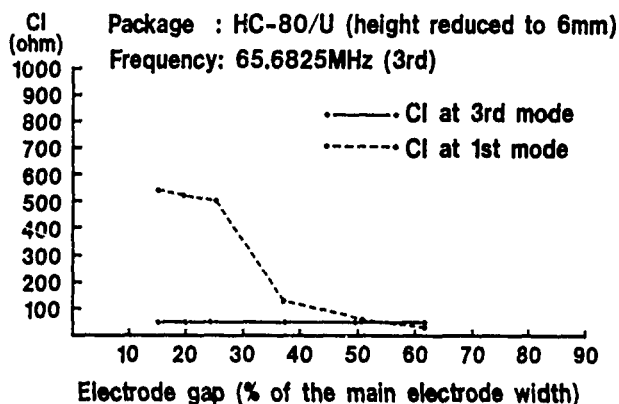


Fig. 6 The relationship between crystal impedance (CI) and electrode gap.

Sample Experiments

Experiment 1

Fig. 7 is the admittance circle diagram for a 65.6825MHz third-order overtone resonator mounted in a 6mm high HC-80/U package. The specifications are shown in Table 1.

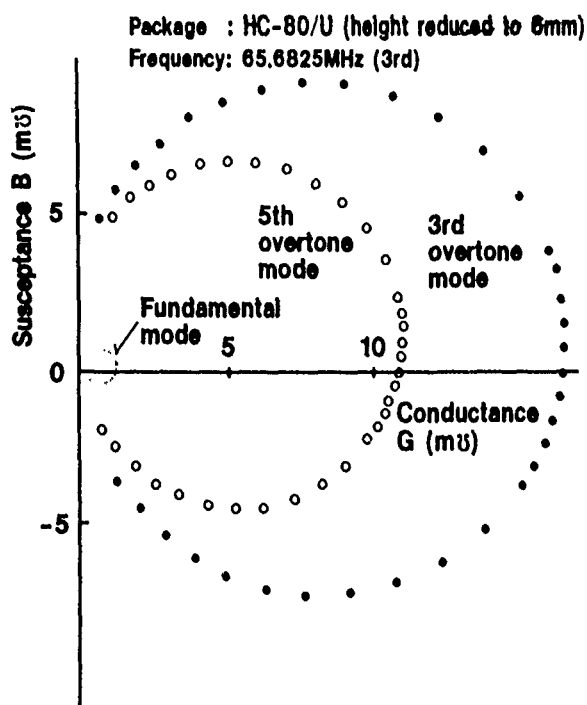


Fig. 7 The admittance circle diagram of an overtone mode assigned AT-cut resonator at the harmonic resonant frequencies.

The equivalent resistance at the desired third-order overtone is 60 ohms, while the resistance at the fundamental frequency is 610 ohms. Due to internal friction loss in the resonator, the resistance at the fifth and higher harmonic modes is substantially larger than at the third harmonic mode. The admittance circle characteristics at the third-order overtone of this resonator are indistinguishable from those of conventional designs.

Table 1 Specifications of a Third-Order Resonator

Package	HC-80/U (height reduced to 6mm)
Electrode width	1.83mm
Electrode gap	0.23mm
Mass loading	410kHz (silver)

Experiment 2

Fig. 8 shows typical data of a resonator designed to minimize the equivalent resistance at the seventh-order overtone mode. The specifications of this resonator are shown in Table 2. To achieve satisfactory performance, the resistance at the seventh, fifth, third and fundamental modes should be inversely proportional to the square of the frequency. The experimental resonator met these conditions, with a slight additional margin at the seventh-order overtone. In most oscillation circuits, the negative resistance has a cutoff frequency below which the negative resistance changes to positive resistance; consequently, commercial resonators need not provide the strict inverse-square attenuation of the lower overtone modes shown in Fig. 8.

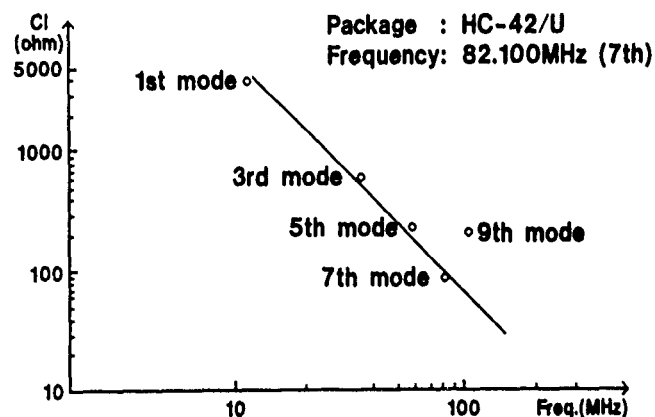


Fig. 8 Crystal impedance at the harmonic resonant frequencies.

Table 2 Specifications of a Seventh-Order Resonator

Package	HC-42/U
Plate width	7.0mm
Electrode width	2.0mm
Electrode gap	0.5mm
Mass loading	1.5kHz (aluminum)

Circuit Applications

When the resistance at the desired overtone is sufficiently low, it is possible to treat these resonators in the same way as fundamental-mode resonators. Because there is no need to supply the frequency selective negative resistance required by previous resonators, it becomes possible to produce coilless, calibration-free overtone-mode oscillation circuits. Here, we will report on practical

applications in high-speed CMOS and transistor-Colpitts oscillator circuits.

Fig. 9 shows a clock-pulse generator/driver (μ PD-71611) for the NEC Type V-60 32-bit microprocessor. The resonator is a third-order assigned mode AT-cut crystal housed in an HC-49/U package. It is designed to operate at 32MHz. The response in the fundamental mode is suppressed.

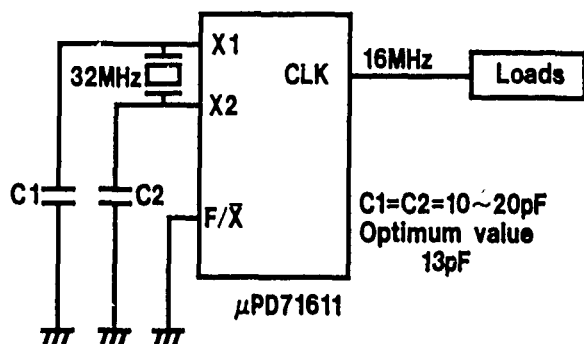


Fig. 9 An application of an overtone mode assigned AT-cut crystal resonator in a high-speed CMOS gate oscillator.

Fig. 10 shows a collector-ground Colpitts oscillator designed to operate at 108MHz. The 108MHz resonator is a fifth-order assigned mode AT-cut crystal housed in an HC-80/U package (height reduced to 6mm). The response in the third and fundamental modes is suppressed.

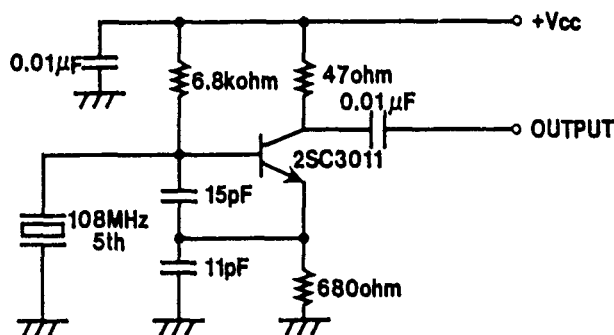


Fig. 10 An application of an overtone mode assigned AT-cut crystal resonator in a Colpitts transistor oscillator.

When we measured the frequency-temperature characteristics of the above resonators, we found that these resonators maintained the same excellent stability as previous AT-cut overtone-mode resonators. These resonators are pin compatible with previous overtone-mode resonators, and we confirmed that they performed satisfactorily in conventional oscillators using LC circuits.

Conclusion

By employing subelectrodes in an AT-cut overtone mode resonator, it is possible to minimize the equivalent resistance at a selected overtone mode. Using such resonators, it is possible to configure high-frequency resonator circuits that are coilless and require no calibration.

Acknowledgements

We would like to acknowledge the assistance of Mr. Tsuguo Ueda of the Semiconductor Applications Technology Division of Nippon Electric Company who suggested a practical application for the new resonator design in the clock pulse generator for the V-60 microprocessor, greatly speeding the analysis process.

We would also like to thank Mr. Yuzo Mochizuki of Shizuoka University, who contributed in the form of constant feedback about the analytic results.

References

- (1) Shockley W., Curran D.R. and Koneval D.J.: "Energy Trapping and Related Studies of Multiple Electrode Filter Crystals", 17th Annual Frequency Control Symposium, pp.88 (1963)
- (2) Onoe M. and Jumonji H.: "Analysis of Piezoelectric Resonators Vibrating in Trapped Energy Modes", J. IECE of Japan, Volume 48, No. 9 pp.1574-1581 (Sep, 1965)
- (3) Mason W.P.: "Electro-mechanical Transducers and Wave Filters", D. Van Nostrand Company, Inc. (1948)
- (4) NEC catalogue No. IP-5410 (Japanese).

Overtone Response of Composite Bulk
Acoustic Resonators

J. Rosenbaum and H. L. Salvo, Jr.
Westinghouse Defense & Electronics Center
Baltimore, MD 21203

S. V. Krishnaswamy
Westinghouse R&D Center
Pittsburgh, PA 15235

Abstract

The performance of miniature composite bulk resonators is usually optimized by designing the passive supporting layer as thin as possible consistent with fabrication and reliability constraints. If overtone operation is considered, the resonator performance (defined as 1/2 the ratio of Q and C_r ratio) can approach that of the edge supported structure for proper ratios of piezo to supporting film thicknesses. The one-dimensional Mason model is employed to study the overtone response of important composite structures including ZnO on Si and AlN on GaAs. Sensitivity of the response with respect to acoustic attenuation and metal mass loading is considered and correlation with measured results is presented.

Introduction

Thin film composite resonators consist of a sputtered layer of piezoelectric zinc oxide or aluminum nitride on top of a supporting substrate, usually silicon or gallium arsenide. Such an arrangement permits the integration of the resonator with active semiconductor devices. In addition the substrate provides mechanical support to the soft piezolayer. Combining the mechanical, acoustic and thermal properties of the films, high performance, low temperature coefficient resonators have been fabricated.^(1,2,3)

For the one dimensional case where the thickness of the resonator is much less than its lateral dimensions an equivalent circuit consisting of a clamped capacitance C_0 in parallel with a series combination of motional resistance R_m , capacitance, C_m and inductance L_m provides valuable insight into the performance of practical devices. Two significant performance characteristics are the resonator C_r ratio:

$$C_r = C_0 / C_m$$

and the Q given by: $Q = \omega_0 L_m / R_m$. Combining them provides a single figure of merit:

$$F.O.M. = Q / 2C_r$$

Acoustic resonators are usually studied by solving the coupled field equations, subject to the relevant electrical and mechanical boundary conditions. In the case of a simple crystal resonator the application of the acoustic boundary conditions result in the elimination of even order harmonics and decreased electromechanical coupling for the odd orders resulting in decreased performance characteristics for higher order overtones especially in filter applications. Because of the complexity of the coupled electrical and mechanical equations, the application of the field equations to composite piezoelectrically active and nonactive layers of arbitrary thicknesses is quite cumbersome. While it is possible to consider the nonactive layer as a perturbation on the piezoelectric film, observation of such devices with relatively thick passive layers shows characteristics which vary considerably from those of a simple freely oscillating piezoelectric resonator.

Lakin and Wang have derived a closed form expression for the electrical impedance of a composite resonator using the one dimensional Mason model.⁽⁴⁾ They showed that a structure of ZnO on Si possesses a resonant response at even order harmonics but since they did not include acoustic loss in their model a determination of the device Q was not possible. While this model provides a qualitative picture of the electromechanical coupling it is not suitable for detailed analysis and design of practical devices. We have extended the Mason model approach to include the effects of acoustic attenuation in the active and passive layers as well as mass loading and lossy metallic layers. This approach lends itself naturally to a composite geometry since additional layers of arbitrary thickness and acoustic attenuation can be used with the original structure without resolving a complex set of equations. The impedance characteristic provides sufficient information to determine the electrical and acoustic performance of the resonator. These include: the resonant frequency, f_r , the values of the motional elements C_m , L_m , R_m , the capacitance ratio, C_r and the clamped capacitance C_0 . The inputs required are the layer thicknesses, the acoustic velocities, attenuations, and

TABLE II

CIRCUIT ELEMENTS

Area = $5 \times 10^{-8} \text{ m}^2$
 ZnO : 3.6 microns
 SiO₂ : 5.0 microns
 $k^2 = .06$

FUNDAMENTAL

MODEL		#407
f_r (GHz)	.308	.303
R(Ohms)	20	350
Q	3780	670
L_m (nH)	32	127
C_m (pF)	.008	.0022
C_0 (pF)	1.2	1.23
C_r	150	560
F.O.M.	12.6	.6

SECOND HARMONIC

f_r (GHz)	.715	.719
R(Ohms)	3	3
Q	2100	2100
L_m (nH)	1.4	1.4
C_m (pF)	.036	.035
C_0 (pF)	1.2	1.23
C_r	32	30
F.O.M.	33	35

Conclusions

1. A one-dimensional model predicts the performance of composite resonators only if energy trapping is present. In the ZnO/Si system this is only true for harmonic operation.
2. The second harmonic performance is within a factor of two of the edge supported structure.
3. The performance peak in harmonic operation is strongly dependent on the normalized thickness ratio as well as the placement of the counter electrode.
4. It is possible to design and fabricate high perform film bulk resonators up to at least 2 GHz using composite structures operating in the second or third harmonics.

References

1. T. Grudowski, J. Black, T. Reeder, D. Cullen and R. Wagner, "Fundamental Mode VHF/UHF Bulk Acoustic Wave Resonators and Filter on Silicon," Proc. 1980 IEEE Ultrasonics Symposium, 1980 pp. 829-833.
2. J. Wang and K. Lakin, "Low-temperature Coefficient Bulk Acoustic Wave Composite Resonators," Appl. Phys. Lett., Vol. 40, pp. 308-310, 1982.
3. K. Nakamura, H. Sasaki, and H. Shimizu, "ZnO/SiO₂ Diaphragm Composite Resonator on a Silicon Wafer," Electron. Lett., Vol. 17, pp. 507-508, 1981.
4. K. Lakin and J. Wang, "UHF Composite Bulk Resonators," in Proc. 1984 Ultrasonics Symposium, 1980, pp. 834-837.
5. R. Moore, J. Haynes and B. McAvoy, "High Overtone Bulk Resonator Stabilized Microwave Sources," in Proc. 1981 Ultrasonics Symposium, 1981, pp. 414-424.
6. H. Tiersten and D. Stevens "An Analysis of Thickness-extensional Trapped Energy Resonant Device Structures With Rectangular Electrodes on the Piezoelectric Thin Film on Silicon Configuration," J. Appl. Phys., Vol. 54, pp. 5893-5910, 1983.

composite structure had a wide peak near a normalized thickness ratio of 1. This is due to a relatively large coupling constant near this value but the size and position (thickness ratio) of the peak are strongly dependent on the substrate attenuation and metallic layer thickness.

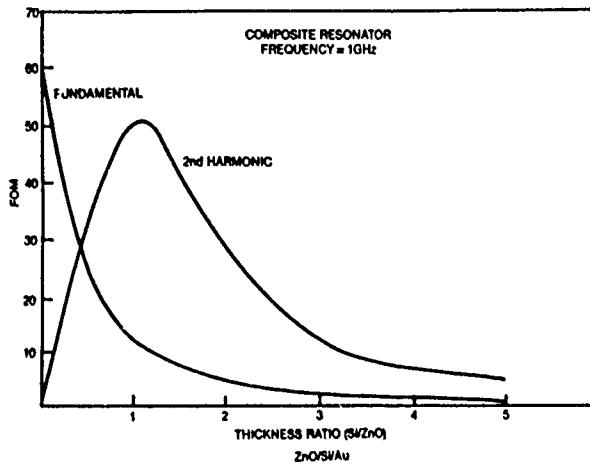


Figure 2. Performance of 1 GHz ZnO/Si composite resonator for 2nd harmonic as compared to fundamental mode.

Figure 3 shows the F.O.M. of a composite resonator at 1 GHz in which the counter electrode is positioned either under the Si or between the Si and ZnO. Note that for the latter geometry (curve b) the performance peak has approximately the same value but is shifted to a thickness ratio of two.

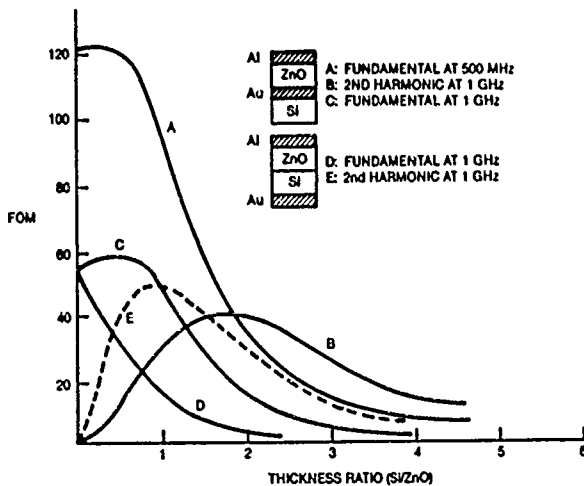


Figure 3. Performance of 1 GHz ZnO/Si composite resonator for 2nd harmonic as compared to fundamental mode.

Figure 4 shows the performance of a 1 GHz resonator for a number of important combinations of piezo and substrate materials. The combination of ZnO on AlN is especially attractive since it combines the high coupling of ZnO with the low attenuation and desirable mechanical properties of AlN (hardness greater than Si). Since the AlN is used as the support layer it need not be piezoactive. Figures 5 and 6 show the performance of third and fourth harmonic composite ZnO/Si structures operating at 1.5 and 2 GHz when compared to the fundamental mode devices operating at the same frequencies. Note that for thickness ratios greater than 1 for the 3rd harmonic and 2 for the fourth harmonic the overtone performance of the composite is higher than the fundamental. At the performance peak of the third harmonic the thicknesses are ZnO:1.7 μ m; Si:4.0 μ m and for the fourth harmonic they are ZnO:1 μ m; Si:5.3 μ m.

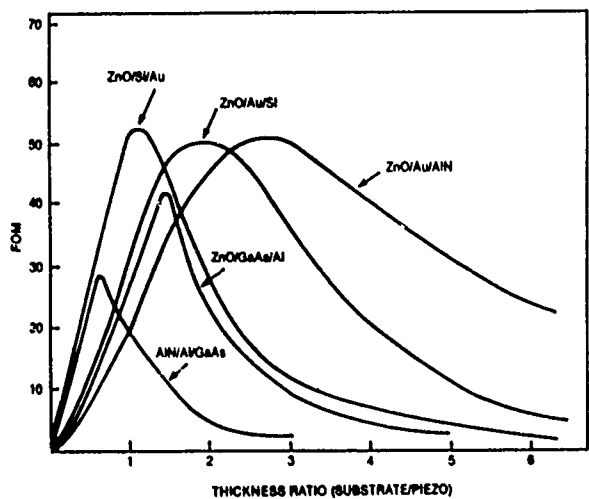


Figure 4. Performance of 1 GHz resonator for a number of important combinations of piezo and substrate materials.

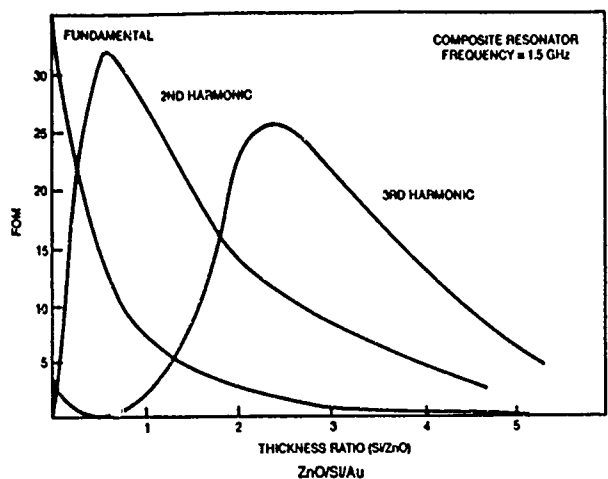


Figure 5. Performance of 1.5 GHz ZnO/Si resonator for fundamental, 2nd and 3rd harmonics.

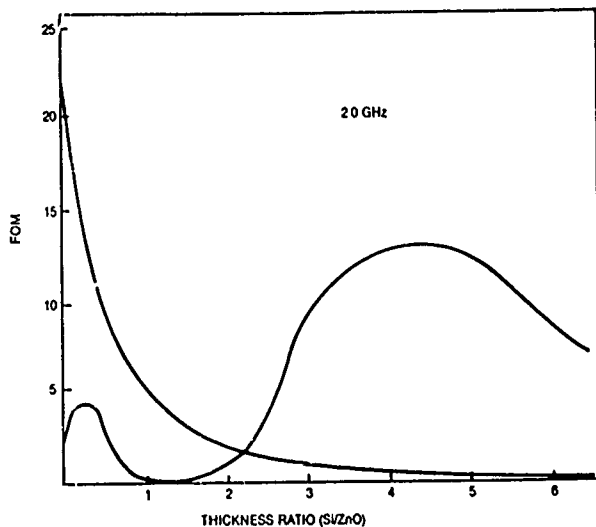


Figure 6. Performance of 2 GHz ZnO/Si resonator comparing fundamental and 4th harmonic modes.

Correlation with Measured Results

Table 1 shows the values of series resonant frequencies for the fundamental and second harmonic modes of a composite resonator calculated using the transmission line model compared to data from typical devices fabricated at Westinghouse. The correlation is quite good in view of the fact that there is no simple relationship between these frequencies even if the mass loading effects of the metallic layers is not included.

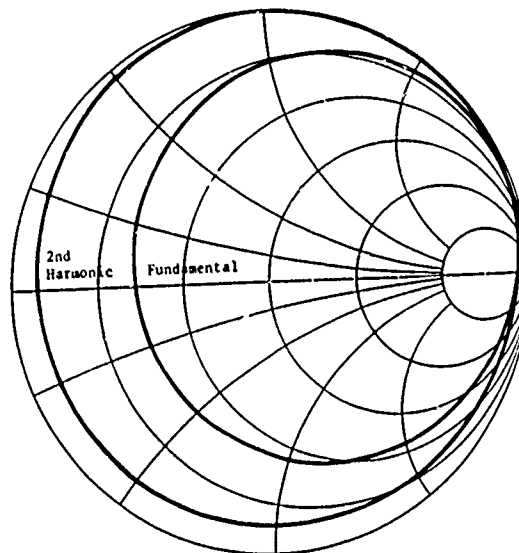
TABLE I

SERIES RESONANT FREQUENCIES OF FUNDAMENTAL AND SECOND HARMONIC

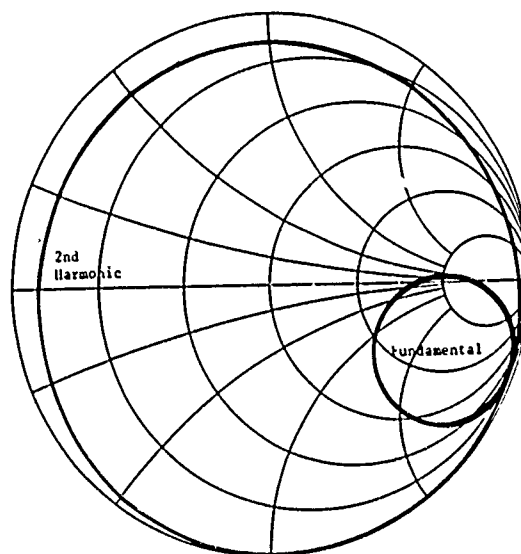
	Fundamental (MHz)	Second Harmonic	Ratio
R401	303	718.5	2.36
R405	302.5	718.8	2.38
R407	297.1	705.5	2.38
Model (Best Fit)	308	715	2.32

Figure 7a and b shows the input impedance characteristic for the model and one of the fabricated devices for the fundamental and second harmonic. The agreement is excellent for the harmonic but extremely poor for the fundamental. This is due to the fact that the ZnO/Si structure is not energy trapped in the fundamental mode which causes a dramatic decrease in the response (increase C_r) even with a high value of k^2 of the ZnO film and a decrease in Q .⁽⁶⁾ This is verified in Table II which shows a comparison of the circuit values for the same devices. Note that both the R_m and C_r are increased in the fabricated

device over those predicted by the model in the fundamental. Had the decrease in Q been due to non uniform etching causing a wedge shape the C_r would not decrease; if the poor fundamental response had been due to poor quality piezofilms the harmonic response would also have been seriously degraded.



(a)



(b)

Figure 7. Impedance characteristic of a) best fit model and b) fabricated device for fundamental and 2nd harmonic.

TABLE II

CIRCUIT ELEMENTS

Area = $5 \times 10^{-8} \text{ m}^2$
 ZnO : 3.6 microns
 SiO₂ : 5.0 microns
 $k^2 = .06$

FUNDAMENTAL

	MODEL	#407
f_r (GHz)	.308	.303
R(Ohms)	20	350
Q	3780	676
L_m (nH)	32	127
C_m (pF)	.008	.0022
C_0 (pF)	1.2	1.23
C_r	150	560
F.O.M.	12.6	.6

SECOND HARMONIC

f_r (GHz)	.715	.719
R(Ohms)	3	3
Q	2100	2100
L_m (nH)	1.4	1.4
C_m (pF)	.036	.035
C_0 (pF)	1.2	1.23
C_r	32	30
F.O.M.	33	35

Conclusions

1. A one-dimensional model predicts the performance of composite resonators only if energy trapping is present. In the ZnO/Si system this is only true for harmonic operation.
2. The second harmonic performance is within a factor of two of the edge supported structure.
3. The performance peak in harmonic operation is strongly dependent on the normalized thickness ratio as well as the placement of the counter electrode.
4. It is possible to design and fabricate high perform film bulk resonators up to at least 2 GHz using composite structures operating in the second or third harmonics.

References

1. T. Grudowski, J. Black, T. Reeder, D. Cullen and R. Wagner, "Fundamental Mode VHF/UHF Bulk Acoustic Wave Resonators and Filter on Silicon," Proc. 1980 IEEE Ultrasonics Symposium, 1980 pp. 829-833.
2. J. Wang and K. Lakin, "Low-temperature Coefficient Bulk Acoustic Wave Composite Resonators," Appl. Phys. Lett., Vol. 40, pp. 308-310, 1982.
3. K. Nakamura, H. Sasaki, and H. Shimizu, "ZnO/SiO₂ Diaphragm Composite Resonator on a Silicon Wafer," Electron. Lett., Vol. 17, pp. 507-508, 1981.
4. K. Lakin and J. Wang, "UHF Composite Bulk Resonators," in Proc. 1984 Ultrasonics Symposium, 1980, pp. 834-837.
5. R. Moore, J. Haynes and B. McAvoy, "High Overtone Bulk Resonator Stabilized Microwave Sources," in Proc. 1981 Ultrasonics Symposium, 1981, pp. 414-424.
6. H. Tiersten and D. Stevens "An Analysis of Thickness-extensional Trapped Energy Resonant Device Structures With Rectangular Electrodes on the Piezoelectric Thin Film on Silicon Configuration," J. Appl. Phys., Vol. 54, pp. 5893-5910, 1983.

A QUARTZ FLUID DENSITY SENSOR PRESSURE TRANSDUCER

Roger W. Ward
Errol P. EerNisse

QUARTZTRONICS, INC.
1020 ATHERTON DRIVE, BLDG. C
SALT LAKE CITY, UT 84123
(801) 266-6994

SUMMARY

A pressure transducer utilizing a quartz crystal tuning fork, similar to the ones used in quartz wrist watches, has been developed. The transducer is based upon the observation that the frequency of the tuning fork is a strong function of the density of the gas surrounding the tines of the tuning fork. Since the density of the gas is a function of both the pressure [1] and the temperature [2],[3] of the gas, either parameter may be sensed if the other is known or controlled.

Since the working gas surrounding the crystal may be chosen from any number of inert gases, it is possible to adapt the frequency sensitivity of the tuning fork, which determines the resolution of the pressure transducer, by selecting an inert gas with the appropriate molecular weight. The frequency sensitivity of the tuning fork to variations in gas density may also be adjusted by varying the tine thickness/width ratio. Finally, using simple design rules, the transducer may be configured in conjunction with the working gas selection and tuning fork design to achieve a variety of full scale measurements.

Several prototypes have been constructed using pistons, metal bellows, and elastomer diaphragms for isolating the gas surrounding the tuning fork from the external pressure medium. They can be used as laboratory instruments or for general purpose transducer applications. Resolution is typically better than 0.2% of full scale. Accuracy and repeatability is typically 0.4% for some designs.

BACKGROUND

The background for the fluid density effect was described by Ward and EerNisse [2]. They applied the fluid density effect to the construction of a filled thermal system thermometer.

Hirata et. al. [4] have described the use of a quartz tuning fork as a vacuum gauge, by utilizing its impedance change as a function of gas pressure. From vacuum to one atmosphere the change in impedance ($Z-Z_0$) changes 4 orders of magnitude. However, for higher accuracy it is advantageous to measure the frequency change--a digital format--instead of the analog format of the

impedance. There are advantages, though, in that the impedance change is relatively independent of crystal contamination, as reported by Hirata.

Since the density of an ideal gas is a function of both its temperature and pressure, the fluid density effect upon a quartz tuning fork has now been applied to the measurement of pressure.

THEORY

It was shown in [2] that the fractional frequency shift, df/f , of a vibrating tuning fork is given by:

$$df/f = C \frac{er}{w\tau^2 q} \quad (1)$$

where C is a constant (nearly 1), e and w are the thickness and width, respectively, of the tuning fork's tines, r is the density of the gas, and $\tau^2 q$ is the density of the tine material.

From (1) it is apparent that the frequency of a vibrating element is substantially linearly dependent upon the density of the working fluid, and that increased sensitivity of a tuning fork can be achieved by both increasing the thickness-to-width ratio of the vibrating tine and selecting a more dense fluid as the working medium.

Experimental verification of equation (1) was given in reference [2].

Considering (1) with regards to: 1) Charles' and Boyles' gas law,

$$\frac{P_0 V_0}{T_0} = \frac{P_1 V_1}{T_1} \quad (2)$$

where P_0 and P_1 , and V_0 and V_1 are the pressures and volumes of a system at temperatures T_0 and T_1 (T in K); and 2) the Universal Gas Law:

$$PV = nRT \quad (3)$$

where n is the number of molecules in the system and R is the Universal Gas Constant ($R = 8.317 \text{ J/mole K}$); it becomes clear that if temperature is known (or constant) and a known, well-behaved (i.e. near ideal) gas is

used, a fluid density sensor (FDS) pressure transducer can be constructed.

However, as explained in [2], the compressibility, Z [5], of real gases prevents them from having linear density changes with increasing pressure, as predicted by (3), and as classically explained by Van der Waal. Thus, the frequency/pressure response of the tuning fork is non-linear over large pressure ranges.

ELASTIC MODULI EFFECTS

Stockbridge [6], in 1965, identified the effect of hydrostatic pressure on the elastic moduli of quartz as being the primary cause of frequency-pressure effects observed on AT- and BT-cut quartz crystal microbalances used for deposition rate monitors in vacuum evaporator systems. This effect would result even if the mass, and therefore the density, of the gas approached zero, so it is not to be confused with the gas density effect which occurs as a result of sound generation in a gas. The pressure induced elastic moduli effect is extremely small (8×10^{-8} /psi for the AT-cut) and is totally obscured by the gas density effect on quartz tuning forks. Note that all thickness shear mode devices (ie, AT and BT cut resonators), have shearing amplitudes on the order of a few angstroms--hardly an efficient "mover" of air. Hence, the gas density effect on thickness shear mode resonators is negligible, whereas tuning forks have relatively large time displacements which result in a large gas density effect.

PRESSURE TRANSDUCER DESIGN

Several transducer design have been constructed and tested. These have used pistons (Figure 1), metal and elastomer diaphragms (Figure 2), and metal bellows (Figure 3).

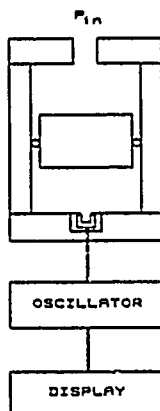


Figure 1: Piston Design FDS

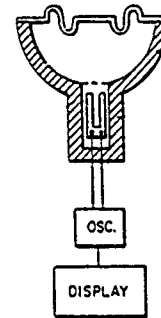


Figure 2: Diaphragm Design FDS

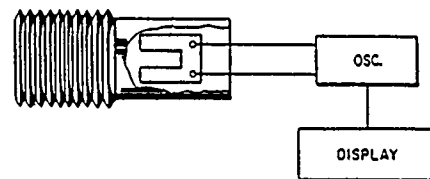


Figure 3: Metal Bellows Design FDS

A FDS pressure transducer requires that gas must always surround the tuning fork; therefore, it is difficult, if not impossible, to provide atmospheric pressure reference to make the FDS a gauge-pressure transducer; hence, the transducer is an absolute pressure gauge. Since the quartz tuning fork is very small ($4 \times 10^{-5} \text{ in}^3 = 6.5 \times 10^{-4} \text{ cm}^3$), it can be mounted into a very small cavity (the internal volume of a 3x8 mm watch crystal can is $3.4 \times 10^{-3} \text{ in}^3 = 5.7 \times 10^{-2} \text{ cm}^3$). Consider Figure 1: if the initial pressure inside and outside the cylinder is 15 psia and the external pressure increases to 30 psia, then the piston will compress the gas inside the cylinder (and therefore the crystal can) to 30 psia, by decreasing the internal volume to (approximately) 1/2 the original volume. Hence, as the external pressure doubles, the internal volume halves (temperature constant), until the piston bottoms out.

Therefore, the upper pressure the gauge is capable of responding to is determined by the ratio of the original cylinder's volume divided by the "dead" volume around the tuning fork; ie, the compression ratio of the gauge.

PRACTICAL DESIGNS

To design a FDS pressure transducer to operate over a selected pressure range requires the consideration of several parameters:

1. Maximum operating pressure

2. Thickness/width ratio of the tuning fork--eq.(2)
3. Density of working gas--eq.(1)
4. Maximum frequency excursion possible for the oscillator/tuning fork
5. Compression ratio of the transducer.

Consider the following examples:

A. Assume a piston gauge (Figure 1) is to be designed with a maximum working pressure of 20,000 psia and a maximum "safe" frequency excursion of 15% = 150,000 ppm. Hence, the maximum permissible sensitivity is 7.5 ppm/psi. Therefore, the combination of thickness/width ratio and the density of the working gas must be chosen to restrict the sensitivity to 7.5 ppm/psi. This dictates either a small value of e/w and/or a low density gas.

With a sensitivity of 7.5 ppm/psi and a frequency resolution of 0.1 ppm, a pressure resolution of $0.1 \text{ ppm} / (7.5 \text{ ppm/psi}) = 0.013 \text{ psi}$ is achieved.

If the gauge is to work down to, say, 10 psia, and up to 20,000 psia, then a compression ratio of 2,000 is required, which is impractical.

A more practical compression ratio is 100. There are three choices: i) backfill to an initial pressure of 200 psia, ii) reduce the maximum operating pressure to 1000 psia, or iii) increase the compression ratio by reducing the dead volume around the crystal, or by designing a larger volume piston/cylinder. Backfilling is a good approach if the application does not call for low pressure performance.

Gas compressibility effects will prevent this 20,000 psi FDS design from achieving the 15% frequency shift predicted from the ideal gas laws. The influence of the compressibility effect becomes appreciable above 3,000 psi for most real gases. Figure 6 shows the actual performance of a piston gauge similar to the one considered in this example A.

Since the FDS is a "gas law" device, it will be strongly influenced by temperature. From (2), if T_0 is 273K (0°C) and T_1 is 274K (1°C), the pressure changes by $1/273$, or about 0.3%. Therefore, the gauge must be either temperature controlled or temperature compensated. As discussed in [2], this can be accomplished by several temperature measurement techniques, but the torsional-mode quartz tuning fork [7,8] is possibly the best since it is small, operates over large temperature ranges, is inexpensive, and operates in a frequency format. To achieve 0.03% error due to temperature effects obviously requires correcting for temperature to an accuracy of 0.1°C; the torsional-mode fork can achieve this accuracy.

B. Assume a diaphragm gauge (Figure 2) is to be designed with a maximum working pressure of 100 psia and a 15% frequency excursion, ie, a maximum sensitivity of 1500 ppm/psi. The maximum photolithographic thickness/width ratio is about $e/w = 1$. This

produces a sensitivity, using nitrogen gas, of 25 ppm/psi. To achieve a sensitivity of 1500 ppm/psi would require a gas density of

$$\frac{1500 \text{ ppm/psi}}{25 \text{ ppm/psi}} = 1.25 \text{ g/l} = 75 \text{ g/l}, \quad (4)$$

which is unrealistic. Radon, the most dense noble gas, has a density of 9.73 g/l. Argon (1.78 g/l) would produce a sensitivity of about 35 ppm/psi, or a total frequency shift of $(100 \text{ psi})(35 \text{ ppm/psi}) = 3500 \text{ ppm} = 0.3\%$.

This isn't really too bad, though. Assuming a frequency stability and resolution of 0.1 ppm, a pressure resolution of $0.1 \text{ ppm} / (35 \text{ ppm/psi}) = 0.003 \text{ psi}$ is achieved.

Again, temperature compensation or control is required.

With a compression ratio of 100, an initial pressure of 1 psia is required--hence, the gauge will work to below atmospheric pressure.

Several possible sensitivity enhancing techniques are being investigated. It appears that a factor of at least 4x improvement may be achievable. This helps at low pressures (example B., above), but is not necessary at pressures above about 4,000 psi ($4,000 \text{ psi} @ 25 \text{ ppm/psi} = 10\%$ frequency shift).

CRYSTAL PARAMETER CHANGES

As might be expected, the impedance of a tuning fork exposed to increasing gas density increases, while its Q decreases. Figure 4 shows this behavior.

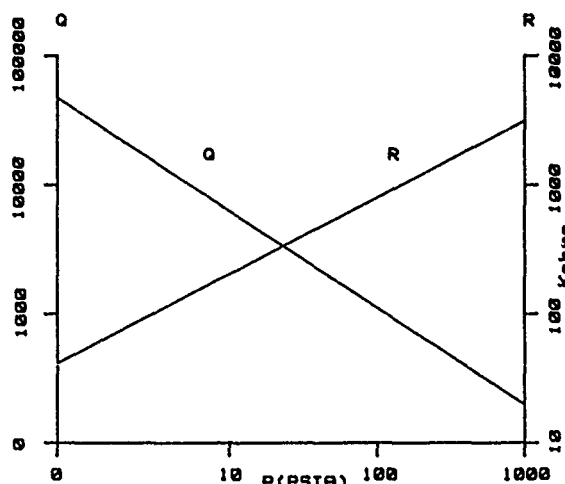


Figure 4: Resistance and Q of a Quartz Tuning Fork vs. Argon Pressure

In order to have good frequency resolution of the FDS, the short-term-stability (STS) of the in-circuit crystal must be considered. The STS is defined as:

$$\left\langle \sigma^2 \frac{\Delta f}{f} \right\rangle = \frac{1}{f_0} \sqrt{\frac{1}{2N} \sum_{i=1}^N (f_{2i} - f_{2i-1})^2} \quad (5)$$

where f_0 is the nominal frequency, f_{2i} is the frequency of the $2i^{\text{th}}$ measurement, f_{2i-1} is the frequency of the $2i-1$ measurement and N is the total number of frequency samples, where the frequency is measured for a specified gate time. The measured STS for a FDS (with oscillator) exposed to Ar gas is shown in Figure 5. It is interesting to note that the STS for any pressure is unchanged for gate-times of one second or less, but it degrades for gate-times greater than 1 second for pressures above 1 atmosphere of argon. This is probably due to the oscillator used for the test contributing more to the STS than the crystal itself.

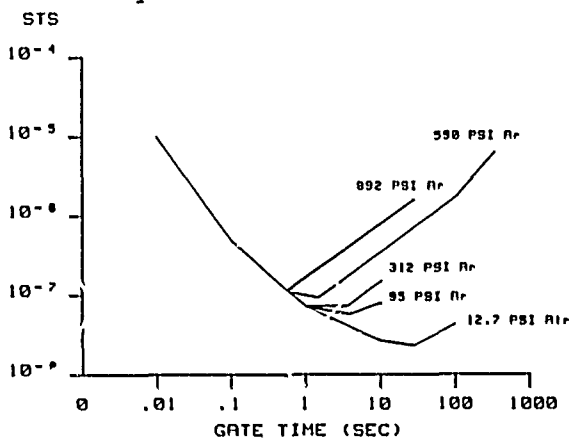


Figure 5: STS vs. Gate Time and Ar pressure

It is also interesting to note that even with a Q of 200 and a resistance of 3 megohms (at 10,000 psi of Ar) the crystal will restart after loss of power, and will exhibit a STS of 2×10^{-7} , which is adequate to resolve 0.003 psi (Example B., above).

RESULTS

Using a piston arrangement similar to Figure 1, the results shown in Figure 6 were obtained.

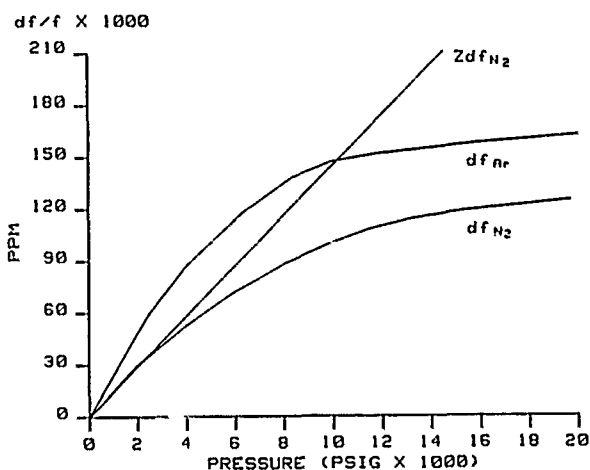


Figure 6: df/f vs. Ar and N_2 Pressure for a Piston Gauge FDS

The reason that the raw data curves shown in Figure 6 are not linear is due to the compressibility effect mentioned above. As can be seen, the result is to substantially reduce the sensitivity of the

gauge at higher pressures. However, the product of $Z df/f$ is a straight line, indicating that the FDS is a true gas density sensor. The compressibility effect also changes the general design rules outlined above, by effectively reducing the average sensitivity of the gauge over the total pressure range.

This piston gauge had a compression ratio of 120. To reach 20,000 psi the piston was backfilled to about 3000 psi with either N_2 or Ar. The Ar back pressure was produced by flowing Ar gas into a pressure chamber immersed in liquid N_2 . This liquified the Ar. A few drops of liquid Ar were then poured into the piston cylinder and sealed.

The main problem encountered in the piston gauge was friction between the piston and the cylinder wall. This caused hysteresis in pressurization/depressurization cycles of 50-100 psi at mid-scale. Zero-return (3000 psi) and end-point (20,000 psi) repeatability was better than 2 psi. Leaving the gauge at ambient pressure for several weeks resulted in no frequency change--i.e., no pressure loss.

Bellows gauges (Figure 2) were also constructed. Repeated attempts and careful study illustrated that compression ratios greater than about 10 cannot be obtained using conventional bellows techniques. Thus, bellows designs have limited pressure range capability; i.e., 10 to 100 psia or 2,000 to 20,000 psia. Over their limited pressure range, the bellows gauges showed excellent resolution and repeatability. One small bellows had a resolution of 8×10^{-3} psi from 10-25 psia.

Diaphragm gauges (Figure 3) with 2 1/2 inch (6.4 cm) elastomer diaphragms were constructed and tested. Performance was excellent with atmospheric backfill to 1000 psia, with a resolution of 0.05 psi and a repeatability of 0.1 psi. With a 40 psia backfill the gauge showed a response to 9000 psi (a compression ratio of 225), but the frequency stability was poor. This was traced to the backfill gas permeating into and through the elastomer. Examination revealed "blisters" to have formed in the elastomer.

CONCLUSIONS

The feasibility of utilizing the effect of gas density on a vibrating quartz tuning fork to measure pressure has been demonstrated. Gauges operating from atmospheric pressure to 20,000 psi have been built and tested. Several problems remain to be solved: friction in piston gauges, limited pressure ranges for bellows gauges, and gas permeation in diaphragm gauges.

The fact that the tuning fork is a true gas law device is well demonstrated by its frequency response versus gas pressure accurately following the compressibility versus pressure curves for the working gases.

REFERENCES

- [1] Ward, R.W., U.S. Patent No. 4,574,639.
- [2] Ward, R.W. and EerNisse, E.P., "A Filled Thermal System Utilizing a Gas Density Sensing Quartz Crystal Tuning Fork", Proceedings, 39th Annual Frequency Control Symposium, U.S. Army Electronics Command, Ft. Monmouth, N.J. and IEEE, 445 Hoes Lane, Piscataway, N.J., 08854 #85CH2186-5, pp. 575-580, 1985.
- [3] Ward, R.W., U.S. Patent No. 4,526,480.
- [4] M. Hirata, K. Kokubun, M. Ono, and K. Nakayama, "Size Effect of a Quartz Oscillator on its Characteristics as a Friction Vacuum Gauge", Vac. Sci. Tech., A 3(3), pp. 1742-1745, 1985.
- [5] Handbook of Tables for Applied Engineering Science, 2nd Ed., Cleveland: CRC Press, 1976, pp. 57-67.
- [6] Behrndt, Klaus H., Editor, Vacuum Microbalance Techniques, Volume 5, Plenum Press, NY, pp. 147-178, 1966.
- [7] R.J. Dinger, "The Torsional Tuning Fork as a Temperature Sensor", Proceedings, 36th Annual Frequency Control Symposium, U.S. Army Electronics Command, Ft. Monmouth N.J., pp. 265-269, 1982. NTIS, 5285 Port Royal Rd., Springfield, VA 22161.
- [8] EerNisse, E.P. and Wiggins, R. B., "A Resonator Temperature Transducer With No Activity Dips", Proceedings, 40th Annual Frequency Control Symposium, U.S. Army Electronics Command, Ft. Monmouth, N.J. and IEEE, 445 Hoes Lane, Piscataway, N.J., 08854, (To be published), 1986.

Copyright 1986 by IEEE.
Presented at 40th Annual
Symposium on Frequency
Control, May 28-30, 1986.
Philadelphia, PA.

A RESONATOR TEMPERATURE TRANSDUCER
WITH NO ACTIVITY DIPS

Errol P. EerNisse and Robert B. Wiggins
QUARTEX, INC.
1020 Atherton Drive, Building C
Salt Lake City, UT 84123

SUMMARY

A thorough experimental and theoretical study has been completed on all four flexural mode families in quartz tuning forks with dimensions similar to torsional tuning forks proposed for use as a temperature sensor. Their interaction with the torsional mode is identified with activity dips, or large increases in torsional mode series resistance, as temperature varies. The experimental and theoretical studies are folded into an empirical set of equations that fit the experimental situation. These equations are used to construct mode interference plots as functions of tuning fork dimensions. A region of dimensions clear of activity dips is identified, which now allows operation of torsional mode sensors over a broad temperature range.

INTRODUCTION

The torsional vibration of a quartz bar or tuning fork tine was examined extensively some years ago.¹ It was shown that at certain crystallographic orientations the derivative of resonant frequency, f_1 , with respect to temperature, T , is zero. This temperature insensitivity is desirable for frequency control; however, for a given crystallographic orientation, the time thickness to width ratio, t/w , plays a strong role in determining the temperature sensitivity. Unacceptable t/w variations in devices manufactured with the present technology limit reproducibility, making the torsional tuning fork an unsatisfactory timebase.

The data collected in the early study clearly showed that the first derivative of f versus T could be large, while the second derivative could be made to become zero for certain t/w ratios and crystallographic orientations. This linear change in frequency with temperature led others to propose the torsional tuning fork as a temperature sensor.²⁻⁴ Its application has been commercialized for temperatures from -50°C up to 125°C . Attempts to use the device above 125°C is fraught with large increases in series resistance, sometimes large enough to terminate oscillation. An example of this is shown in Figure 1. The temperature at which such effects occurred depended on the location of the glue joint in our experiments, which led us to conclude that activity dips, produced by spurious vibrational modes, were the cause of the problem.

FREQUENCY vs TEMPERATURE

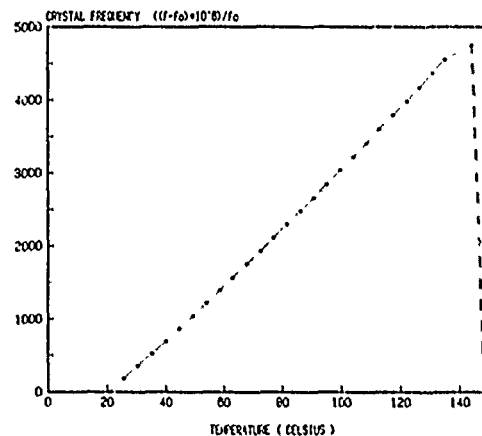


Figure 1. Frequency of a torsional mode device vs temperature showing cessation of oscillation above 125°C due to an activity dip.

This paper reports on an extensive study of the various modes in a quartz torsional tuning fork and the dependences of the different modes on the dimensions. The results allowed empirical relationships to be formulated that are based in part on simple bending beam theory. The empirical formulations allow definition of ranges of dimensions where the torsional mode of the tuning fork can operate over wide temperature ranges free of significant activity dips.

THE MODES OF A FORK

A tuning fork configuration is shown in Figure 2 along with the definition of dimensions L (overall length from fixed end to tine end), m (tine length), t (thickness), w (tine width), and g (gap between tines).

One variation of the configuration in Figure 2 includes a widened support region between the crotch and the glue joint (fixed end), with the increased width being approximately $0.5g$ ($.5 \times$ the gap width) on each side. This modified base region improves mechanical Q in flexural tuning forks, but has not been found to significantly alter the various results to be reported here, so this feature is not treated in detail further. Figure 3 shows a variation that we prefer wherein a widened end portion has been added to help delineate the dimension L independent of the glue joint location, thus lending reproducibility to our

results. We will focus on the fork design in Figure 3 for our discussions.

QUARTZ TUNING FORK

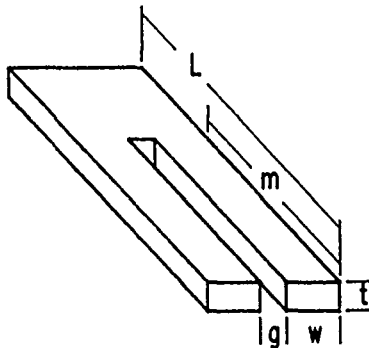


Figure 2. Torsional tuning fork and dimension definitions.

QUARTZ TUNING FORK WITH WIDE ATTACHMENT

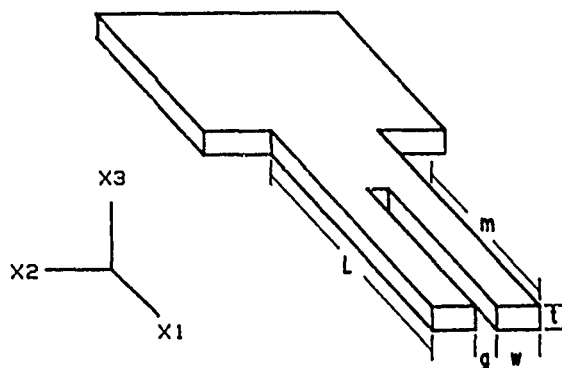


Figure 3. Torsional tuning fork with widened end portion for fixing L. Coordinate system shown also.

There are four flexural families of modes in addition to the torsional mode. The first mode family we call the Pitch Mode. The tines move in unison out of the major plane of the device. Figure 4 shows the second mode of the Pitch family. Displacement occurs in the support region (between the fixed end and the tines) even in the first mode, so we can deduce that L and t are predominant dimensions.

The second family is the Walk mode. Here, the tines also move out of the major plane of the device, but the two tines move out-of-phase. Figure 5 shows the second mode of the Walk family. Here, m and t seem to be the predominant dimensions, but because the support region twists, L has a minor effect. The L dependence is less important in the Walk mode than in the Pitch mode because the support region is stiffer in twisting than in the bending seen in the Pitch modes.

The third family we call the B family; the tines move together in the major plane of the device. Figure 6 shows the first mode of the B family. Here, it is difficult to generalize the dimensional dependences. For the lower modes, the support region is stiff and m and w predominate. For the higher modes L becomes more important than m.

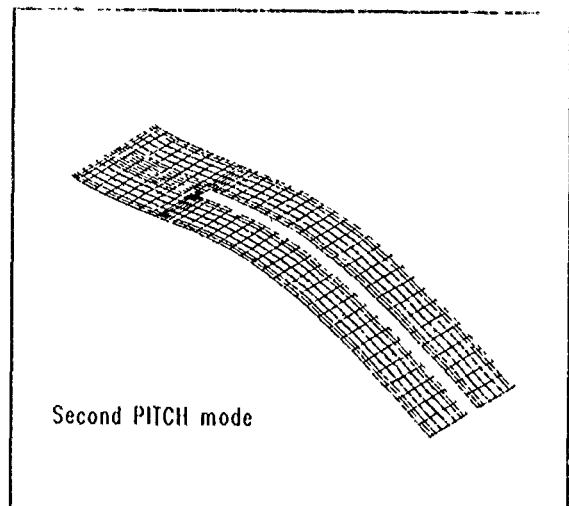


Figure 4. Deflection shape for the second mode of the Pitch family.

The fourth family we call the A mode, where the tines move out of phase in the major plane of the device. This mode family is the usual flexural tuning fork family, which has been studied extensively for timekeeping purposes. In the torsion tuning fork, these modes are simply in the way. Figure 7 shows the first mode of the A family.

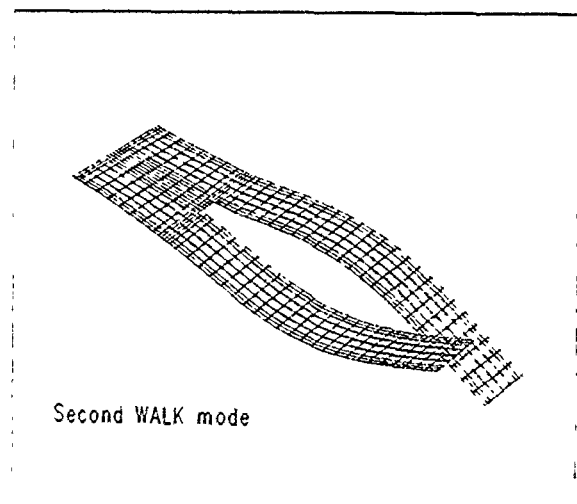


Figure 5. Deflection shape for the second mode of the Walk family.

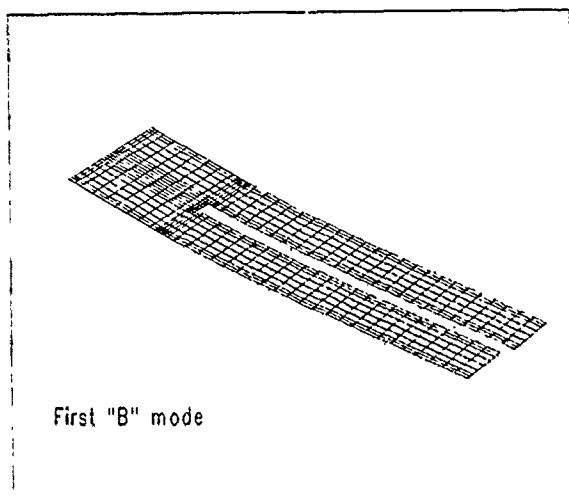


Figure 6. Deflection shape for the first mode of the B family.

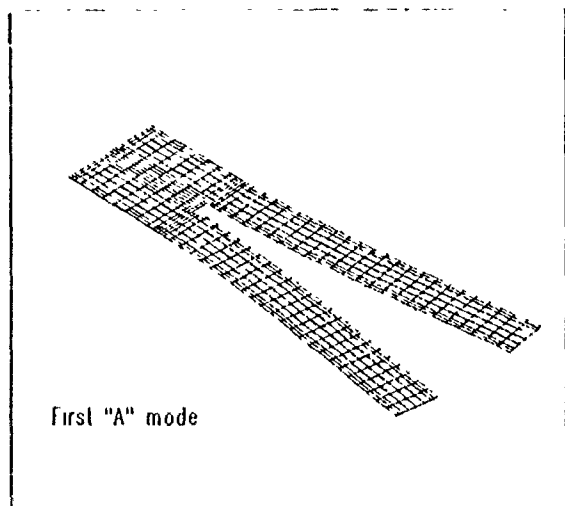


Figure 7. Deflection shape for the first mode of the A family.

For almost all crystallographic orientations, the temperature coefficient (first derivative of f with T) of all four families is negative, particularly for temperatures above 100 C. In contrast, (since the torsional mode is typically made of z cut quartz for ease of photolithographic etching) the torsional mode temperature coefficient is positive. The relative rate at which the torsional mode frequency moves towards the flexural families as temperature increases can be 50 to 75 ppm/C. This means that if there is a flexural mode within 1% of the torsional mode, an activity dip can occur before T changes by as much as 140 to 200 C.

A 1 or 2% window may not seem difficult to achieve. Consider, however, that the torsional mode frequency typically occurs above the 3rd or 4th members of the Pitch and

Walk families and above the 2nd or 3rd member of the A and B families, so the spectrum of modes is quite dense. The probability for the torsional frequency f_t to equal the frequency of one of the flexural modes at some temperature is high, and becomes higher if one desires larger temperature ranges. Consider also the fact that a sensor, like a time base, is expected to perform to 1 ppm or better in frequency if it is to be a precision device. Errors of this magnitude may be produced by nonlinear effects within the quartz which cause coupling between modes with integer ratios of frequency. Thus, although a smaller effect, activity dips wherein the torsional frequency f_t is twice one of the flexural mode frequencies can diminish the sensor's performance. Combining the case where f_t is equal to a flexural frequency with the case where f_t is equal to twice a flexural frequency, leads to a situation wherein the mode spectrum is extremely dense. It becomes possible to design a torsional tuning fork that has an unintentional activity dip within the desired temperature range or, what is more difficult to avoid, to have a marginal design wherein the length, L (normally fixed by the glue joint), has to be held so tightly in production that it becomes impractical.

In order to emphasize the point, we plot the torsional frequency in Figure 8 on top of a mode spectrum for 1x and 2x the various flexural mode frequencies for a typical 262 kHz fork as determined in our work. Note the density of the spectrum around the torsional mode. The density would be even higher if we included the situation of 0.5x the flexural families in Figure 8. We would be able to study in that case the activity dip situation where at some temperature $f_t = 0.5x$ a flexural family frequency. This would require us to treat the four mode families up to the 7th or 8th member, as well as bring into play the extensional mode family where the device lengthens along L and m . We have been unable to study the members of the flexural families that high up either experimentally or theoretically with any confidence, so we treat only 1x and 2x here.

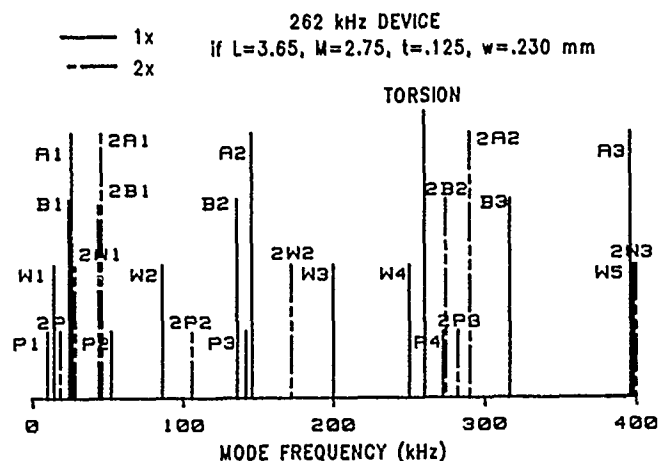


Figure 8. Frequency spectrum plot (mode strength for illustration purposes only) for the 262 kHz device for $L/m = 1.33$. Note the density of modes near the torsional mode.

THEORETICAL ANALYSIS

Finite element analysis (FEA) was used to establish the general mode shapes as well as relative frequencies within a given mode family and between mode families. The mode shapes in Figures 4-7 were obtained with FEA where the entire device was modeled. Since it was necessary to treat many modes, and since all FEA software programs have run time and numerical accuracy problems as the product of degrees of freedom (number of nodes in the model times the degrees of freedom at a node) times the number of mode frequencies (eigenvalues) to be calculated increase, we performed most calculations by dividing the families by symmetry. In this way, we could model only one-half the device, which reduced the number of nodes by one-half and doubled the number of resonant frequencies which we could calculate. The line of symmetry is the center line of the device. The Pitch family was treated by invoking zero rotation about L on the center line of the support region. The Walk family was treated by invoking zero deflection out of the device plane on the center line of the support region. In both cases, no deflection in the w direction was allowed on the center line in the support region. Use of plate bending elements with membrane stresses led to results for the desired modes. Since the boundary conditions for the Pitch and Walk modes did not eliminate the A mode family, its frequencies were solved for in each analysis.

The B family required a 2-d model where the full device was modeled. In this case we obtained the A family as well. Results are too extensive to show here. Suffice it to say that the comparison to the experiments in the following section was good enough to accomplish several things. We are confident that we did not miss any families in our definition of families nor did we miss any members in a given family when we did the experiments. In a few instances where we could not experimentally see a particular mode frequency with confidence, we filled it in by interpolating with the FEA model results.

Incidentally, the torsional mode comes out in the FEA with the symmetry of the Pitch family. It is shown in Figure 9. The support region curves significantly but predominate dimensions for determining the frequency are m, t, and w.

The quantitative accuracy of FEA becomes less for the higher modes (15-20% difference between experiment and theory), as expected. It was therefore necessary to obtain real-world experimental data in order to treat the activity dip problem with accuracy sufficient to establish design rules.

EXPERIMENTAL METHODS AND RESULTS

All four families under investigation are flexural. Simple bending beam theory tells us that all four families depend

primarily on the same elastic constant, which is the compliance along the m (x_1) direction, or the inverse of compliance, Young's modulus. Secondary effects will arise from Poisson's ratio effects as some modes begin to look like plate modes if t/w gets too small, and from a second Young's modulus if bending across w becomes significant. One problem experimentally is to change orientation in order to piezoelectrically excite all four families without changing the elastic constant that controls the different families. The XY plane of quartz offers sufficient variety of piezoelectric coupling while providing the same Young's Modulus for any direction in the plane.

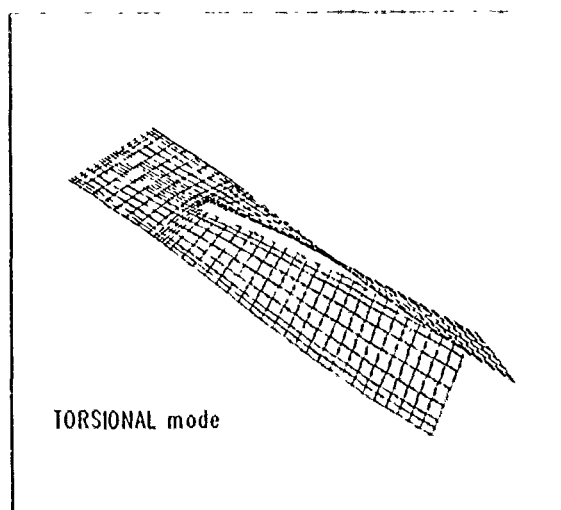


Figure 9. Deflection shape for the torsional mode.

The results from such an experiment can be used for orientations out of the XY plane, because, again, all four flexural families depend (to first order) on the same Young's modulus. The only change is in the frequency of the torsional mode relative to the frequencies of the four flexural families, which will be taken care of in our empirical model by using the appropriate shear compliances controlling the torsional mode relative to the appropriate Young's Modulus controlling the four flexural families.

The Walk mode was excited by using the conventional electrode pattern used for the torsional mode, except with the tines aligned along the Y crystal axis. This provided coupling via the d_{12} piezoelectric constant.

The Pitch mode was excited with a pattern similar to the one used for the Walk mode except the polarity of the two tines was made the same to drive the two tines in unison instead of out-of-phase as in the Walk modes.

The A modes were driven with the usual electrode pattern for a photolithographically-produced watch crystal with the tines along the Y axis, again using the d_{12} piezoelectric constant.

The B modes were driven with a modification of the A mode electrode pattern wherein the tines, along the Y axis, were driven in-phase instead of out-of-phase.

The torsional mode was driven conventionally with the tines along the X direction where the d_{26} piezoelectric constant could be used.

In all four flexural families, the ideal piezoelectric coupling would be obtained by a complex set of electrodes that reversed polarity along L and m at the zeros of curvature. Since we are only interested in identifying the frequencies, not in optimizing the coupling, we chose the simplest possible pattern with no polarity changes along m. A fairly short length for the electrode pattern was used (approximately 30% of m) to excite as many of the overtones as possible.

Frequencies were determined from a pi network circuit with a spectrum analyzer. The basic dimension w was chosen as 0.023 cm, with g = 0.013 cm, and m = 0.30 cm. The L/m ratio was varied by creating different photolith masks. The t/w ratio was varied by using wafers of different thicknesses. Data from three samples of each type were taken to confirm repeatability.

Wafers of different thickness were used to vary t/w from 0.33 to 0.55. The A and B families did not depend on t/w. The Pitch and Walk families were linearly dependent upon t/w, as simple bending beam theory would predict. The torsional mode depended slightly less than linearly on t/w, as theory predicts.

TABLE I
EXPERIMENTAL AND FEA PREDICTED
MODE FREQUENCIES

L = 4.216 m = 2.972 t = .076 w = .230 mm

Mode	Experimental (kHz)	Predicted (kHz)
P1	*	4
W1	7	8
B1	18	21
A1	19	23
P2	25	24
W2	42	43
P3	*	65
W3	92	100
B2	96	111
W4	**	135
A2	121	139
P4	123	125
T	171	204
P5	*	205
B3	229	260
W5	232	255
A3	329	380
P6	333	301

* Mode not observed: electrode pattern drove only even numbered modes.

** Higher order modes difficult to detect due to low Q.

Table 1 shows the frequencies for the various modes as predicted by the FEA compared to the frequencies measured experimentally. Table 1 also shows an example of our interpolation: the odd numbered Pitch modes were not observed experimentally, but the agreement between the observed and FEA predicted frequencies for the even numbered pitch modes allowed us to interpolate the odd numbered modes with confidence.

EMPIRICAL MODELS

While the FEA provides us with the trend in f with dimension changes for an empirical model, the experimental data can be used to quantify an empirical model. We have some guidelines for constructing empirical models from simple beam and plate theory. The A family is the closest to a simple beam because the out-of-phase motion keeps the support region quiescent. The beam theory says that

$$f_{AN} = A_{AN} w k/m^2 \quad (1)$$

where

$$k = (E/\rho)^{1/2} \quad (2)$$

E is Young's Modulus, and ρ is mass density. The constants A_{AN} as determined by our experimental fit are found in Table II. They can be compared to the classical beam solution values of 3.52, 22.03, and 61.70. The differences are due to rotational inertia and shear effects that are neglected in beam bending theory, and to slight errors in our dimensions because of faceting during the quartz etching. No attempt was made to bring the A_{AN} any closer to the results of simple beam theory because our intent is to ratio all the frequencies. It will be the relative values of the A constants between members of the same family and between families that will determine the location of activity dips. The model in Eq. 1 conforms to our experimental observation that L and t do not come into the relationship.

TABLE II
Experimentally Determined Values For
 A_{AN} , A_{BN} , A_{PN} , and A_{WN}

$A_{A1} = 3.05$	$A_{B1} = 8.57$
$A_{A2} = 18.99$	$A_{B2} = 41.82$
$A_{A3} = 51.92$	$A_{B3} = 96.97$
$A_{P1} = 3.91$	$A_{W1} = 4.96$
$A_{P2} = 22.12$	$A_{W2} = 30.91$
$A_{P3} = 59.09$	$A_{W3} = 71.82$
$A_{P4} = 114.1$	$A_{W4} = 89.80$
$A_{P5} = 189.9$	$A_{W5} = 143.4$
$A_{P6} = 281.8$	$A_{W6} = 180.8$
	$A_{W7} = 286.9$

The B family has very little dependence on L/m for the first mode, but becomes strongly dependent for the higher modes. Since we are interested in any B mode that comes near to F_t or $1/2 f_t$, we must model f_{B2} to f_{B4} the most carefully. We find that the following works over the ranges investigated experimentally:

$$f_{BN} = A_{BN} wk/[L^2 (1+m/L)^{1/2}]. \quad (3)$$

The A_{BN} values obtained from our fit to experimental data are found in Table II. The f_{B1} fit is poor; however, we do not need it for this study. We see that Eq. 3 agrees with our experimental observation that t does not come into the relation and that there is a strong L dependence (higher modes in particular).

The Pitch family is flexural in the L and t plane, so we should use t instead of w in the bending formulas. As with the B mode, it is difficult to decide whether L or m should be used. Since the support region comes into play more for the higher order modes near f_t or $0.5 f_t$, we choose

$$f_{PN} = A_{PN} tk/L^2 \quad (4)$$

The values for A_{PN} obtained from a fit to our data are found in Table II. Equation 4 fits our experimental observation of a linear dependence on t and a strong dependence on L.

The Walk family was the most difficult to model and the most difficult to observe experimentally. The FEA errors in comparison to experiment were largest for the Walk family. It varied the most from any attempts to use beam theory because the support region went from very little contribution for the first mode to a fairly strong contribution for the higher order modes. It was thus difficult to pick L or m as the predominant length dimension in a bending formula. After some trial and error, the following seemed to be adequate for the higher modes near f_t or $0.5 f_t$.

$$F_{WN} = A_{WN} t k/[m^2(1+L/m)^{1/2}] \quad (5)$$

Where the A_{WN} values obtained from a fit to our experimental data are found in Table II. Equation 5 fits our experimental observation that the frequency is linear with t and depends on L/m. Of all the families, this one bears further investigation for better understanding.

Finally, the torsional mode frequency has been analyzed in terms of an infinite series. This was deemed awkward for our purposes. A variational calculus model was developed using the x_i coordinates in Figure 3 that led to the much simpler formula

$$f_t = Z t/(w m CD) \quad (6)$$

where

$$C = (1 + t^2/w^2)^{1/2} \quad (7)$$

$$D = (1 + S_{55} t^2/S_{66} w^2)^{1/2} \quad (8)$$

$$Z = G/(\rho S_{66})^{1/2} \quad (9)$$

and $G = 0.45$ for a clamped/free beam. Our Eq. 6 gives the same quantitative answer as the infinite series model¹ to within 1% over the entire range of t/w values investigated. It is a simpler form for finding the temperature behavior of f_t if differentiation with respect to any of the material coefficients or the dimensions is needed.

The Equations 1, 3, 4, and 5 fit the A, B, Pitch, and Walk experimental data to within 5% over the L/m range of 1.3 to 1.5 and t/w range of 0.3 to 0.55. Use of the empirical model beyond the range of dimensions studied is not recommended.

MODE INTERFERENCES

The purpose of formulating empirical models was to avoid extensive experimental data taking. If one chooses to use experimental data to find what combinations of L, m, t, and w cause a mode interference, the data matrix has to be extensive. Fortunately, the linear dependence on t for the Pitch and Walk families, and no t dependence for the A or B families, made our modeling simpler. Since we are interested in avoiding situations where f_t equals 1x or 2x another mode frequency we can most readily define where these relations do occur, and then stay away from them. We thus form

$$f_t/f_{AN} = [Z/(kA_{AN})][(tm)/(w^2 CD)], \quad (10)$$

$$f_t/f_{BN} = [Z/(kA_{BN})][(tL^2)(1+m/L)^{1/2}]/[w^2 mCD], \quad (11)$$

$$f_t/f_{PN} = [Z/(kA_{PN})][L/(m w CD)], \quad (12)$$

and

$$f_t/f_{WN} = [Z/(kA_{WN})][m(1+L/m)^{1/2}]/[wCD] \quad (13)$$

We can set the ratios in Eqs. 10 to 13 equal to 1 or 2, depending on the type of activity dip to be studied. We choose the m/w vs t/w plane to present results. Figure 10 shows a mode interference plot for frequency ratio 1 and L/m = 1.45. The open regions in Figure 10 are the desired choices for dimensions in order to avoid f_t equaling another mode frequency.

In Figure 11, we show a mode interference plot for frequency ratio = 2 and L/m = 1.45. The open regions in Figure 11 are the desired choices for dimensions in order to avoid f_t equaling twice another mode frequency.

Finally, in Figure 12, we combine Figures 10 and 11 to show a mode interference plot with ratios 1 and 2 for L/m = 1.45. We see one clear region, around m/w = 13 and t/w

= 0.35, that is most attractive because the sensitivity, or slope of f vs T , is larger for smaller t/w ratios. The other open regions might work, but the larger t/w ratios mean less sensitivity.

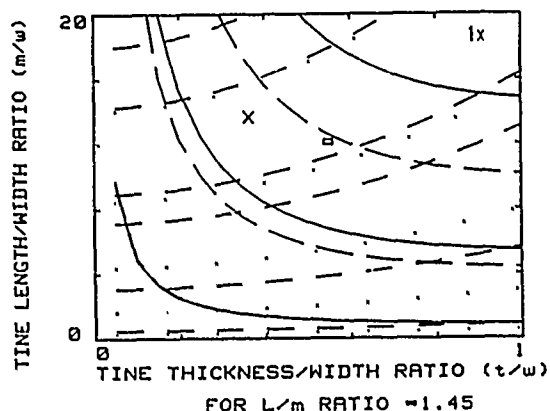


Figure 10. Mode interference plot in the m/w vs t/w plane for $L/m = 1.45$ for when $f_t = 1x$ interfering mode frequency. Square is 262 kHz device. Solid lines descending from upper left are A modes; dashed lines paralleling them are B modes. Dashed lines ascending to upper right are Walk modes; small dots paralleling them are Pitch modes.

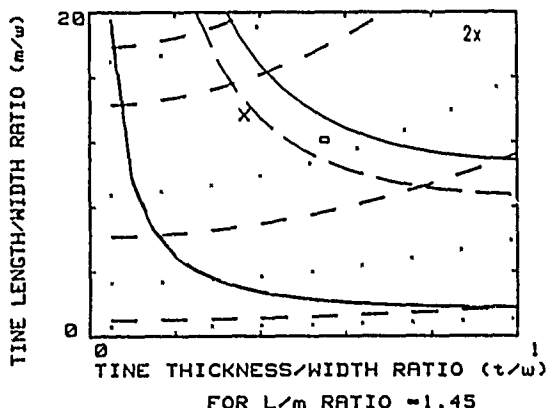


Figure 11. Mode interference plot in the m/w vs. t/w plane for $L/m = 1.45$ for when $f_t = 2x$ interfering mode frequency. Square is 262 kHz device.

Figure 13 shows a mode interference plot like Figure 12 except $L/m = 1.35$. The desirable region, marked with an x, is still clear, so there is not a severe sensitivity to L if one chooses m/w and t/w carefully.

The 262 kHz device² has $m = 0.27$ cm, $w = 0.023$ cm and $t = 0.013$ cm, with a $m/w = 11.9$ and $t/w = 0.56$. We show it in Figures 10 to 13 as a square. We see in Figures 10 to 13 that there are close modes for all three choices of L/m . The interfering mode is apparently, depending on L , f_{p4} or f_{w4} , where f_t equals the interfering mode frequency. For the case where f_t equals $2x$, the interfering mode frequency, depending on L , can be f_{B2} , f_{A2} , or f_{p3} . Thus, the 262 kHz tuning fork is sensitive to L .

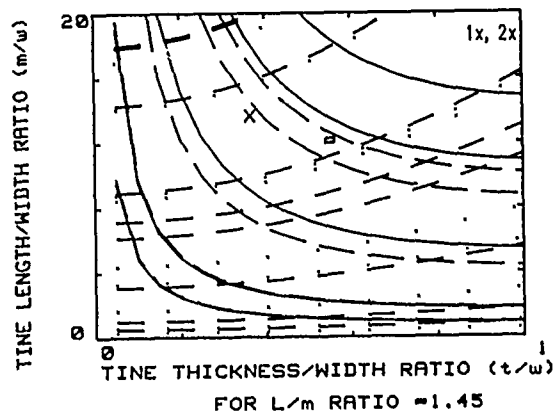


Figure 12. Mode interference plot in the m/w vs. t/w plane for $L/m = 1.45$ for when $f_t = 1x$ and $2x$ interfering mode frequency. Square is 262 kHz device.

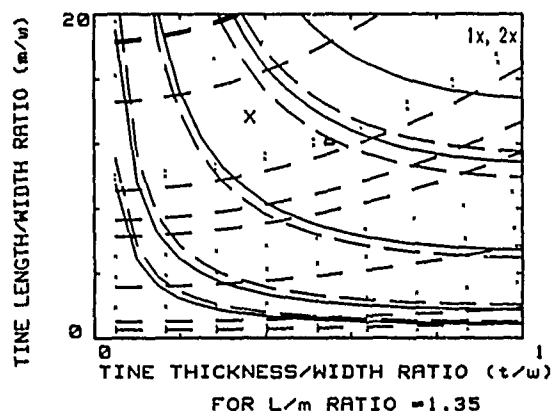


Figure 13. Mode interference plot for $L/m = 1.35$ for when $f_t = 1x$ and $2x$ interfering mode frequency. Square is 262 kHz device. The x marks the recommended region.

CONCLUSIONS

This study has identified the relevant families of modes in a tuning fork structure that could interfere with the torsional mode. Extensive experimental results and theoretical modeling led to the development of empirical formulas relating fork dimensions to frequencies. Armed with the empirical formulas, it is possible to construct mode interference plots showing what choices of dimensions to avoid. Based on these mode interference plots, design choices can be made to avoid activity dips and make the design somewhat insensitive to the location of the glue joint used to fix the device to a substrate.

As an example, Figure 14 shows the frequency spectrum for a device near the x in Figure 13. Note the clear area around the frequency of the torsional mode. Also, Figure 15 shows the residual error in $^{\circ}C$ after a

third order fit of experimental data taken from -50° to 200° C of a torsional tuning fork temperature sensor of dimensions $L = 0.401$ cm, $m = 0.297$, $w = 0.023$ cm, and $t = 0.094$ mm. Frequency was nominally 170 kHz. Note the smooth fit over the entire range with no indication of activity dips. We have investigated several design choices in the desired range; all designs had results similar to Figure 15.

REFERENCES

1. J. Hermann, "DT - cut Torsional Resonators," Proceedings, 31st Annual Symposium on Frequency Control, pp. 55-61 (1977).
2. R. J. Dinger, "The Torsional Tuning Fork As A Temperature Sensor," Proceedings, 36th Annual Symposium on Frequency Control, pp. 265-269 (1982).
3. R. J. Dinger, et al., U.S. Patent 4,437,773.
4. H. Ziegler, "A Low Cost Digital Temperature Sensor," Sensors and Actuators 5, pp. 169-178 (1984).
5. Y. Tomikawa, et al., "A Quartz Crystal Tuning Fork With Modified Basewidth for a High Quality Factor: Finite Element Analysis and Experiments," IEEE Transactions on Sonics and Ultrasonics SU-20, pp. 217-223 (1982).
6. Robert D. Blevins Formulas For Natural Frequency And Mode Shape, p. 108, Van Nostrand Reinhold Company, 1979.
7. Errol P. EerNisse and Robert B. Wiggins, U.S. Patent pending.

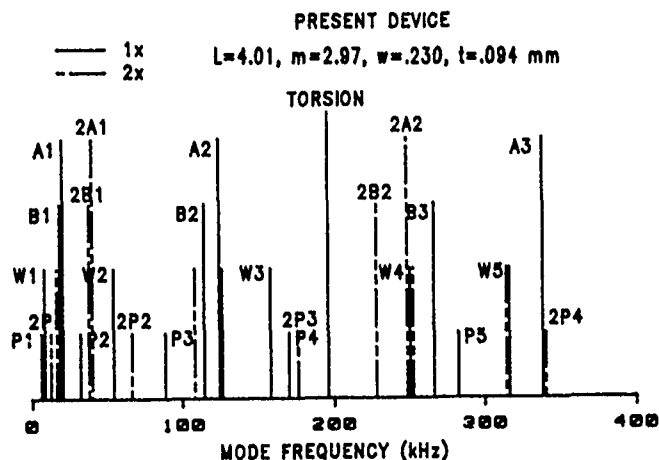


Figure 14. Frequency spectrum plot (mode strength for illustration purposes only) for a device at the x in Figure 13. Note the clear area around the torsional mode.

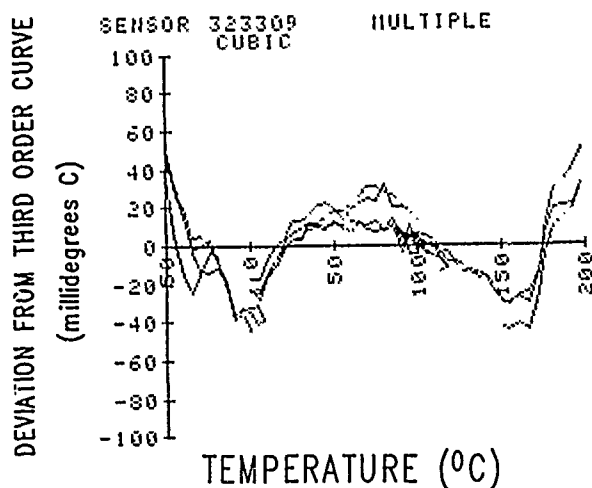


Figure 15. Residual after third order least-squared-error fit of T versus f_c .

TEMPERATURE SENSOR USING QUARTZ TUNING FORK RESONATOR

Toshitsugu Ueda, Fusao Kohsaka, Toshio Iino, Daisuke Yamazaki,

Yokogawa Hokushin Electric Corporation, Corporate R&D
Musashino-shi, Tokyo, JapanSummary

This paper describes the development of a temperature sensor based on a temperature-sensitive quartz "tuning fork" resonator. The resonator was fabricated using photolithography and anisotropic etching. The optimum cut of the crystal and optimum resonator orientation on the wafer were determined -- taking into account the anisotropic etching properties of the crystal -- to give a large temperature coefficient of frequency consistent with small equivalent series resistance.

The tuning fork temperature sensor operates over a wide range of temperature (from 4.2K to 250°C). After the temperature sensor was cycled between 0°C and 4.2K, hysteresis in the temperature characteristic was only 0.005K at 0°C. The quartz temperature sensor is quite compact -- it is mounted in a hermetic case only 3 mm in diameter.

Introduction

Thermometers which use quartz resonators as temperature sensors have good temperature resolution, and their output -- i.e. frequency -- is easy to process digitally¹. But conventional quartz thermometers using thickness-shear mode resonators have some weak points: they are physically large, and have narrow temperature-measuring range (from -80°C to 250°C) and large hysteresis.

The object of our research was to reduce hysteresis by using a "tuning-fork" shaped quartz resonator made by photolithography and anisotropic etching, and to increase the temperature span -- in particular, to enable temperature measurements down to the temperature of liquid helium, 4.2K (i.e. a span from 4.2K to 250°C). Compared with mechanical machining, chemical etching -- the fabrication process we adopted -- lends itself well to mass production and to produce arbitrary shapes readily without surface damage^{2,3}.

Because quartz crystal has anisotropic properties, we must determine the conditions for first-order temperature coefficient of frequency (TCF) to be sufficiently large for it to be used as a thermometer sensor element. Anisotropy of the piezoelectric strain coefficients also affects the equivalent series resistance. The etching properties of quartz crystal are also anisotropic, so we must also determine the condition of low equivalent series resistance and how best to fabricate it by chemical etching. This paper discusses the theory, experimental results, and characteristics as temperature sensor.

PrinciplesResonant Frequency and Its Temperature Characteristic

Because quartz crystal has anisotropic properties, the resonant frequency of a quartz resonator depends on the angle of cut, ϕ , and orientation (angle ψ on the wafer) as shown in Fig. 1. We fabricated resonators with various orientations on wafers of various angles of cut, and studied the resultant resonator characteristics. (The definitions of the coordinate system involved -- "angle of cut" and "element orientation angle", were chosen from manufacturing considerations). By applying Newton's laws of motion to vibrating beam, and taking into account the effects of rotational inertia and shearing deformation, solving the resultant differential equations, we obtained the following expression for resonant frequency:

$$f = \frac{\alpha^2}{4\pi \sqrt{3\rho s'_{22}}} \frac{w}{\ell^2} \frac{1}{\sqrt{1 + \frac{\alpha^2 N}{12} \left(\frac{w}{\ell}\right)^2 \left(1 + \frac{s'_{44}}{k' s'_{22}}\right)}} \quad (1)$$

f : resonant frequency

α : root of $\cos(\alpha) \cdot \cosh(\alpha) + 1 = 0$

$N = - \frac{2 \tanh(\alpha) \cdot \tan(\alpha) + \alpha \cdot (\tanh(\alpha) + \tan(\alpha))}{\alpha \cdot (\tanh(\alpha) - \tan(\alpha))}$

ρ : density of quartz

s'_{22}, s'_{44} : elastic compliance constants
after coordinate transformation

k' : constant of rigidity (≈ 0.85)

w : width of resonator

ℓ : length of resonator

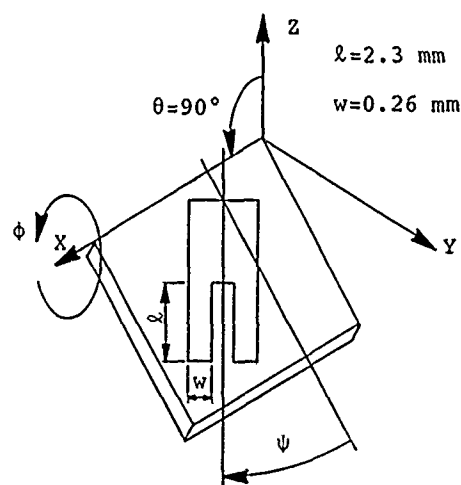


Fig. 1 Orientation of a Resonator on a Quartz Wafer.
(Directions of Arrows Correspond to Positive)

First order temperature coefficient of frequency (TCF) is expressed as a function of linear expansion coefficients and temperature coefficients of elastic compliance of quartz in equation (2).

$$\frac{1}{f} \cdot \frac{df}{dT} = -\frac{1}{2}(k_1 \alpha_{\perp} + k_2 \alpha_{//} + \alpha'_{22} + \alpha'_{44}) \quad (2)$$

α_{\perp} : linear expansion coefficient perpendicular to Z axis
 $\alpha_{//}$: linear expansion coefficient parallel to Z axis
 $\alpha'_{22}, \alpha'_{44}$: first order TCF of s'_{22}, s'_{44}
 k_1, k_2 : constants determined by direction cosines

Figure 2 shows contour plot of first order TCF at 0°C calculated with ϕ and ψ as parameters. We used Bechmann's constants for the temperature coefficients of elastic stiffnesses⁴.

For use as a temperature sensor, a resonator with large first order TCF is required — so we must choose suitable fabrication angles ϕ and ψ . It is impracticable to realize a resonator having linear frequency temperature characteristics over a wide temperature range; it is much more practical to fabricate a resonator with large TCF over a wide temperature range, and use a microprocessor to linearize the characteristic. Taking into account etching properties and equivalent series resistance, we decided to fabricate a resonator which was aligned parallel to the Y axis on a wafer cut parallel to the X axis at an angle of -40.23° to the Y axis. The theoretical value of the TCF at 0°C is -58.2 ppm/K for first order TCF, -26.9×10^{-3} ppm/K² for second order TCF, and -2.8×10^{-6} ppm/K³ for the third order TCF.

Spread in Temperature Coefficient of Frequency

As shown in equation (2), TCF is determined by linear expansion coefficient and temperature coefficients of elastic compliance constants, angles θ , ϕ , and ψ , and dimensions l and w (see Fig. 1). Variations in angles and dimensions due to fabrication errors cause deviations in TCF. If we design a resonator such that TCF is not affected by fabrication error, then high-precision thermometers can be realized with fewer temperature calibration points, because all resonators have nearly-equal first order TCFs. Δ TCF, the deviation in first order TCF, can be evaluated from equation (3) as the total TCF deviation caused by deviation of each parameter:

$$\Delta \text{TCF} = \left| \frac{\partial \text{TCF}}{\partial \theta} \right| \cdot \Delta \theta + \left| \frac{\partial \text{TCF}}{\partial \phi} \right| \cdot \Delta \phi + \left| \frac{\partial \text{TCF}}{\partial \psi} \right| \cdot \Delta \psi \quad (3)$$

Fabrication precision of angles — the precision when wafers are cut from quartz blanks, and the precision of alignment between wafer and photomask — is better than 0.5° . Dimensional precision — determined by the precision of photomask and precision of control of etching — is of the order of several μm . We found that an angular error of 0.5° had the same effect on TCF as a dimensional error as large as $42 \mu\text{m}$ — so dimensional fabrication errors are negligible compared with angular errors. Figure 3 shows a contour plot of the deviation in first order TCF (Δ TCF), assuming that all angles can have fabrication error of 0.5° . Figure 3 shows that, for a resonator defined by $\phi = -40.23^\circ$ and $\psi = 0^\circ$, the deviation in first order TCF (due to fabrication error) is less than 0.8 ppm/K or 1.3% of first order TCF.

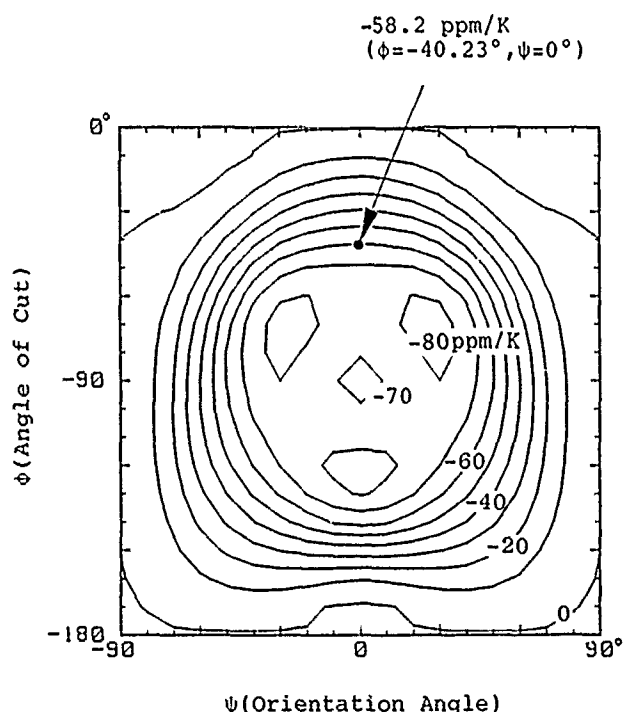


Fig. 2 Contour Plot of First Order TCF vs. Angle of Cut ϕ and Orientation Angle ψ .

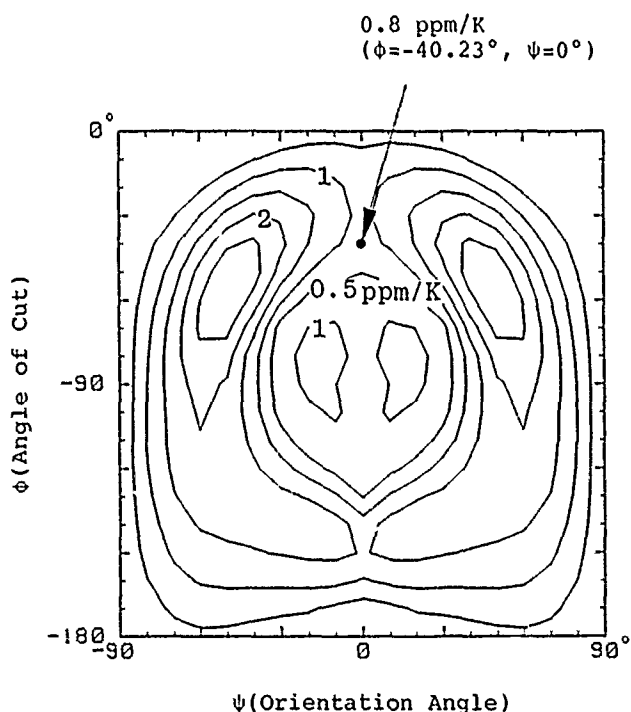


Fig. 3 Contour Plot of Deviation in First Order TCF Caused by Fabrication Errors vs. Angle of Cut ϕ and Orientation Angle ψ . (Angular Errors of 0.5°)

Etching Properties

The following characteristics are desirable for fabricating quartz resonators by chemical etching:

- (1) fast etching rate
- (2) little edge under-cutting
- (3) relatively regular-shaped etched faces

The polar diagram in Fig. 4 shows the results of the experiment — the measured etching rates, for wafers cut at fifteen different angles of cut ϕ , with spline function interpolation, and the magnitude of calculated and measured first order TCF at 0°C for resonators oriented at $\psi = 0^\circ$. Wafers were etched in saturated $\text{NH}_4\text{F} \cdot \text{HF}$ solution at $82 \pm 2^\circ\text{C}$.

To satisfy condition (2), we should choose an angle of cut ϕ such that the ratio A/B is large — where A is the etching rate perpendicular to the wafer surface, and B is the etching rate perpendicular to A . In Fig. 4, the wafer cut at $\phi = -40.23^\circ$, Miller index (0223), satisfies conditions (1) and (2), and the first order TCF of a resonator oriented at $\psi = 0^\circ$ on this wafer is as large as -58.2 ppm/K .

Figure 5 shows a SEM cross-sectional photo of the resonator. Its shape is regular, and almost rectangular. A resonator defined by orientation $\psi = 0^\circ$ on a wafer cut at angle $\phi = -40.23^\circ$ showed relatively regular etched faces.

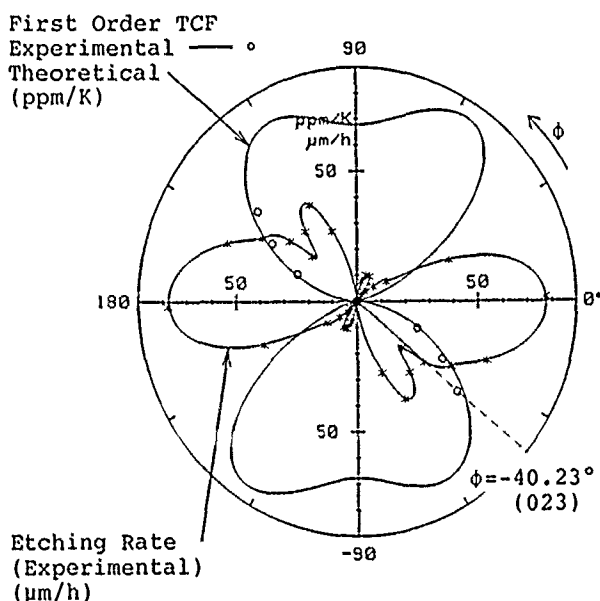


Fig. 4 Polar Diagram of Normal Etching Rate and Absolute Value of First Order TCF. ($\psi = 0^\circ$)

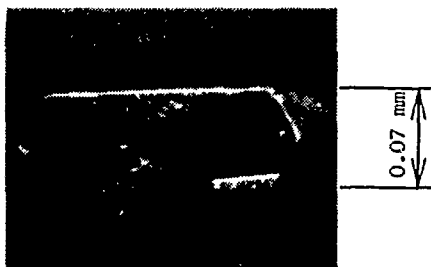


Fig. 5 Cross Sectional View of a Tine of Resonator ($\phi = -40.23^\circ$, $\psi = 0^\circ$).

Equivalent Series Resistance

It is desirable that the equivalent series resistance of the resonator be as low as possible, as the sensor probe cable the resonator is connected to represents a relatively large capacitance load. Equivalent series resistance (ESR) is a function of angle of cut ϕ and orientation ψ , because quartz is anisotropic. The relation between ESR and Q is as follows:

$$Q = \frac{\omega \cdot L}{R} \quad (4)$$

- ω : resonant angular frequency
 L : equivalent inductance
 R : equivalent series resistance
 Q : quality factor

We found that for a given size and shape of resonator element, Q and ω — hence the ratio L/R — were virtually constant, regardless of angle of cut or orientation angle. From equation (4), the Q factor and ratio is virtually constant for resonators with the same size and resonant frequency hence relation of ESR among resonators are as follows:

$$R = R' \cdot \frac{L}{L'} \quad (5)$$

- R : ESR of a resonator
 L : equivalent inductance of a resonator
 R' : ESR of the other resonator
 L' : equivalent inductance of the other resonator

Solving the piezoelectricity equations gives the equivalent inductance L of a tuning fork resonator as follows:

$$L \propto \left(\frac{S'_{122}}{d'_{32}} \right)^2 \quad (6)$$

- S'_{122} : elastic compliance constants
 d'_{32} : piezoelectric strain coefficients

Given the piezoelectric strain coefficients and elastic compliance constants (we used Bechmann's constants), the relative value of L (hence the relative value of R) for a given angle of cut and orientation angle can be calculated by coordinate transformation. We used a cut angle $\phi = -40.23^\circ$ and orientation $\psi = 0^\circ$ as the reference.

Figure 6 shows how the reciprocal of R varies with orientation ψ for an angle of cut angle $\phi = -40.23^\circ$. Figure 6 shows that $1/R$ has a local maximum for $\psi = 0^\circ$, corresponding to a local maximum in first order TCF. This combination (low equivalent series resistance, and high first order TCF) is ideal for a temperature sensor.

Fabrication Process

Figure 7 illustrates the process of fabricating a resonator from a quartz wafer.

- (1) Sputtering Au/Cr thin film on quartz wafer
- (2) Photolithography, etching to define resonator pattern
- (3) Photolithography to define electrode pattern
- (4) Quartz etching in saturated $\text{NH}_4\text{F} \cdot \text{HF}$ soln. at $82 \pm 2^\circ\text{C}$
- (5) Etching of Au/Cr to fabricate electrode

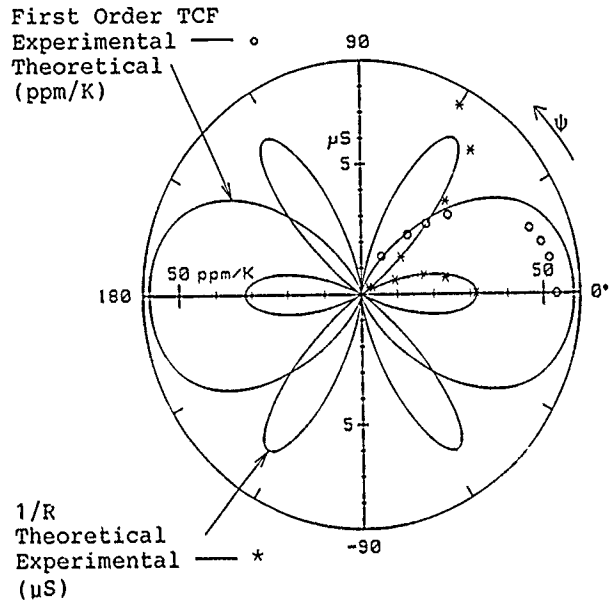


Fig. 6 Polar Diagram of Equivalent Series Resistance and Absolute Value of First Order TCF.
($\phi = -40.23^\circ$)

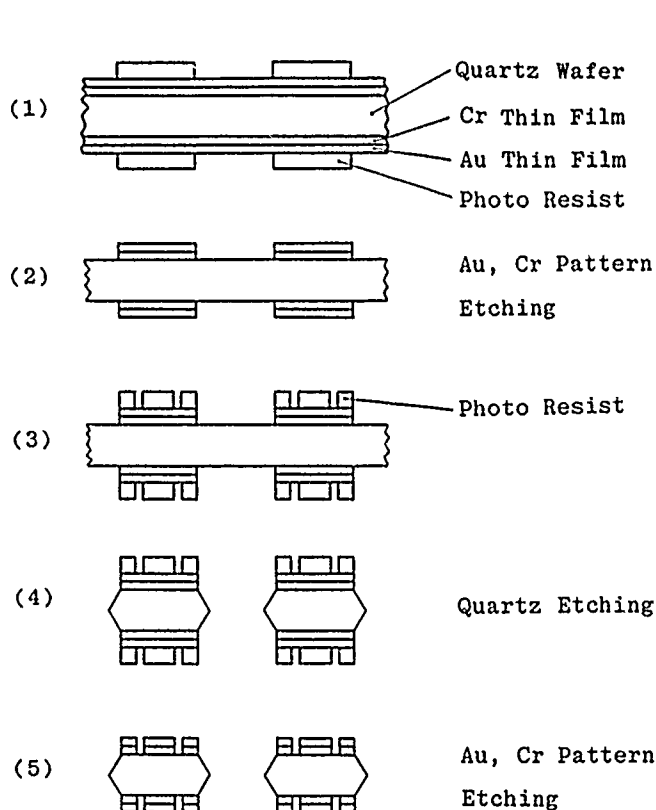


Fig. 7 Fabrication Process of Quartz Resonator.

Shape of a Resonator

Because stresses at the root of each vibrating beam of a tuning fork resonator cancel each other, virtually no energy from tuning fork vibration is transmitted to the resonator support. Tuning fork resonators have the advantage that resonant frequency is virtually unaffected by the way the resonator is supported — and thermal stresses caused by the difference in linear expansion coefficients between quartz and the support material will not affect the resonator frequency — unlike thickness-shear mode resonators, which are difficult to support without resonant frequency being affected.

Although the tuning fork resonator is relatively stress-free, we made a narrow neck between the resonator support and vibrating beams to further reduce the stress transmitted from the support to the vibrating beams (to reduce the effects of thermal stress and thermal hysteresis). We used the finite element method to analyze the resonator and its neck/support, to determine the best shape for them. We calculated the stress in the support, and the stress transmitted to the vibrating beams of the resonator, for a compressive force applied to the support. Figure 8 shows the results. The narrow neck reduces the transmission of stress from support to vibrating beams. The narrower the neck, the less stress transmitted to the beams. On the other hand, a reasonable neck width is desirable, for mechanical strength. Finite element analysis shows that using a neck whose width w_n in Fig. 8 is 35% of the width w_0 of the resonator reduces the stress to one seventh that without a neck.

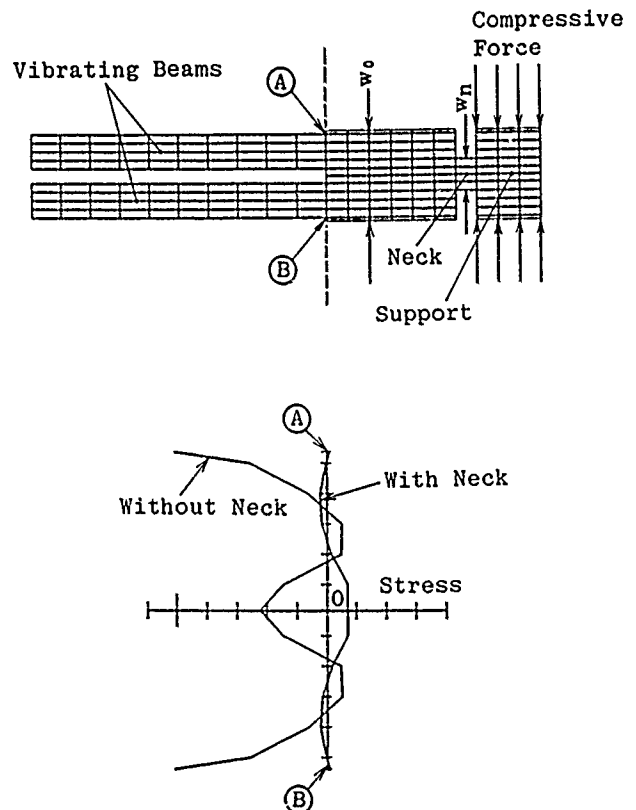


Fig. 8 Stress Calculated with Finite Element Method at the Root of Vibrating Beams When Compressive Force is Applied to Both Sides of Tuning Fork Support.

Construction of Thermometer Probe

Figure 9 shows the prototype of a thermometer probe. Thin film electrodes on the resonator support are soldered with AuSi eutectic solder to two terminals, and the resonator was hermetically sealed in a stainless steel casing by laser-welding. To improve aging characteristics and thermal response, the casing was filled with helium gas. The time constant of thermal response from room temperature to hot water was 0.9 second. To avoid problems caused by changes in cable capacitance (in parallel with the resonator) with cable movement, the cable was enclosed in a stainless steel sheath packed tight with Alumina powder.

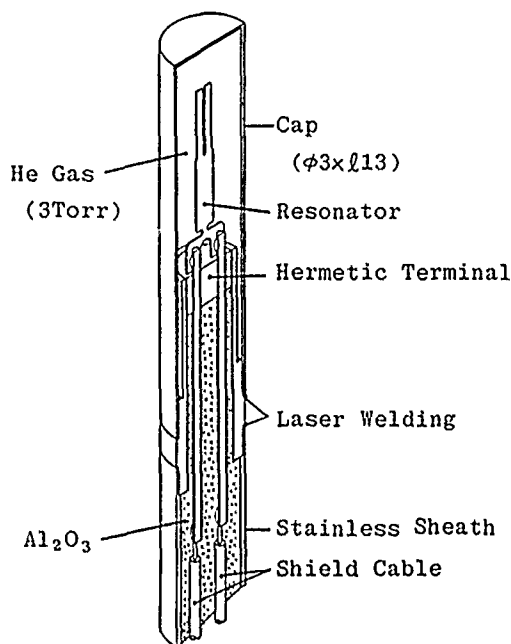


Fig. 9 Construction of Thermometer Probe

Experimental Results

Results of Temperature Tests

The temperature characteristics of the resonator -- its resonant frequency and insertion loss -- were measured between 4.2K (liquid helium temperature) and 250°C and showed in Fig. 10, using the equipment listed in table 1. The characteristics of the resonator at 4.2K and 0°C are shown in table 2. The measured value of first order TCF at 0°C agreed well with the theoretical value. The first order TCF measured at 4.2K, however, was quite different from the theoretical value. One reason, we think, is that Bechmann's and Aruga's elastic stiffness constants obtained at 20°C, and temperature coefficients of stiffness obtained at 25°C (20°C for Aruga's constants), as well as linear expansion coefficients obtained at 25°C were not valid at temperatures far from room temperature^{4,5}. The discrepancy at 4.2K is particularly large, so we must investigate the changes in these constants at low temperatures.

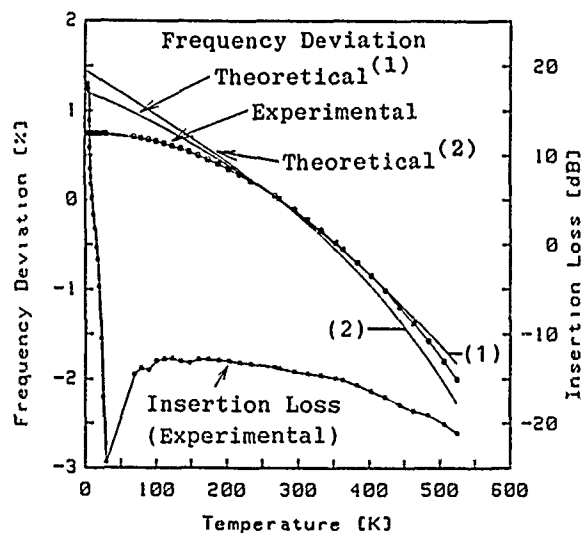
Table 1 Equipment for Temperature Test

Temperature Range	4.2K to 40K	40K to 0°C	0°C to 250°C
Equipment	Cu Block with Heater in Liquid Helium Dewar	Cryogenic Refrigerator	Oil Bath Rosemount Model 910
Standard Thermometer	PtCo Cryogenic RTD YEW Type J263	Pt RTD Rosemount Model 162CE	
Accuracy of Measurement	Less than 0.2K	Less than 0.1K	Less than 0.02K

Table 2 Characteristics of Resonator

Temperature	4.2K	0°C
First Order TCF		
Experimental	-0.34 ppm/K	-54.1 ppm/K
Theoretical(1)	-31.6 ppm/K	-58.2 ppm/K
Theoretical(2)	-50.2 ppm/K	-63.3 ppm/K
Insertion Loss	+18.0 dB	-13.8 dB
Q factor	1,100,000	52,000
Hysteresis(3)	less than 0.05K	less than 0.005K
Hysteresis(4)	less than 0.09K	less than 0.015K
Resonant Frequency	40 kHz	40 kHz
Spread of First order TCF		
Experimental		0.37 ppm/K
Theoretical		less than 0.8 ppm/K

- (1) Bechmann's Constants
(2) Aruga's Constants
(3) Necked Resonator
(4) Conventional Resonator



- (1) Bechmann's Constants
(2) Aruga's Constants

Fig. 10 Temperature Characteristics of Quartz Resonator.

Figure 11 shows the results of temperature calibration between 4.2K and 250°C. The measured temperature vs. frequency characteristic was fitted by a third order spline function as follows:

$$T(f) = A + B \cdot f + \sum_{i=1}^{n-2} b_i \cdot N_i(f) \quad (7)$$

$$N_i(f) = (f - f_i)^3$$

T: temperature

A, B, b_i : constants

f: resonant frequency

f_i : resonant frequency at calibration point

For this third order spline function, the calibration error was less than ± 0.5 K for thirteen temperature calibration points. Accuracy can be improved by using more calibration points.

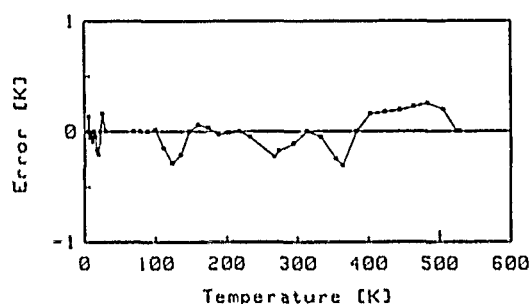


Fig. 11 Calibration Error of Quartz Thermometer

Figure 10 also shows the effect of temperature on resonator insertion loss. Insertion loss D can be defined as:

$$D = \left| \frac{v_o}{v_i} \right| \quad (8)$$

v_o : output signal level of a resonator

v_i : input signal level to a resonator

Insertion loss is proportional to the reciprocal number of Q factor. Figure 10 shows that insertion loss varies greatly with temperature. The reason for the increase in insertion loss around 40K is thought to be impurities in the quartz. Table 2 shows that first order TCF at 4.2K is about 1/160 that at 0°C. On the other hand, insertion loss decreases 31.8 dB representing a 20 times increase in Q factor. If we define the "quality" of a resonator by the product of Q factor and first order TCF, at 4.2K this product is only one eighth of that at 0°C. Thus the resonator operates well even at temperatures as low as 4.2K.

We measured repeatability (hysteresis) of the temperature sensor by thermal cycling between 4.2K and 0°C. Table 2 shows that by using a narrow-necked resonator, repeatability is improved. At 0°C, in particular, the hysteresis (cf. Table 2) is one tenth that of a thickness-shear mode resonator.

Spread in TCF

The maximum spread in first order TCF of resonators fabricated on different wafers was estimated as less than 0.8 ppm/K. To confirm this, we performed temperature tests between 0°C and 80°C for ten resonators made from four different wafers, and calculated their first order TCFs at 0°C. Three times the standard deviation of first order TCF was less than 0.37 ppm/K. This small spread between resonators means that the accuracy of these temperature sensors can be guaranteed by calibration at only a few points.

Conclusions

We studied quartz etching properties and the temperature coefficient of frequency of quartz elements to develop compact temperature sensors with wide temperature range. Temperature tests confirmed that our thermometer was accurate to better than ± 0.5 K over the temperature range 4.2K to 250°C, and hysteresis was negligible.

Future work will involve:

- (1) improving calibration accuracy to further improve accuracy of the thermometer,
- (2) studying the relation between number of calibration points and accuracy, and improving the thermometer probe design.

Acknowledgements

We would like to thank Mr. Tamura and Mr. Harada of Yokogawa Hokushin Electric Corporation for their helpful advice and encouragement during this work, and we would also like to thank Associate Professor Nakazawa of Shinshu University and Dr. Arthur Ballato of the U.S. Army Laboratory Command Electronics Technology and Devices Laboratory for their advice about constants of quartz crystal.

References

- (1) D. L. Hammond, A. Benjaminson, "The Linear Quartz Thermometer — a New Tool for Measuring Absolute and Difference Temperatures", Hewlett-Packard Journal, vol. 16 No. 7, March 1965, pp.1-7.
- (2) T. Ueda, F. Kohsaka, T. Iino, D. Yamazaki, "Quartz Crystal Micromechanical Devices", Digest of Technical Papers, Transducer '85, pp.113-116.
- (3) "Special Issue on 3-D Semiconductor Devices Structures", IEEE Trans. Electron Devices, ED-25, 1978.
- (4) R. Bechmann, A. D. Ballato, T. J. Lukaszek, "Higher-Order Temperature Coefficients of the Elastic Stiffnesses and Compliances of Alpha Quartz", Proc. of I.R.E., vol. 50, August 1962, pp.1812-1822.
- (5) M. Aruga, "The Elastic Constants of Quartz and their Temperature Characteristics", Bulletin of Tokyo Institute of Technology, vol. A-2, December 1956, pp.88-182 (in Japanese).

DOUBLE-ENDED TUNING FORK QUARTZ ACCELEROMETER*

William J. Kass and Gary S. Snow
Sandia National Laboratories
Albuquerque, New Mexico 87115

SUMMARY

Quartz accelerometers with frequency output have been designed and built using etched double-ended tuning fork technology. These devices consist of two tuning forks connected in a cantilever beam configuration with beryllium-copper spacers which serve as a proof mass and inboard spacer. The frequency outputs of each tuning fork resonant circuit are mixed to yield an output frequency of approximately 1kHz at zero acceleration. The sensitivity of the devices is approximately 1Hz/G. The fundamental vibration frequency of the cantilever beam is 2200Hz in this undamped device. Shock testing up to 1500G has indicated good survivability characteristics.

Individual fork measurements compared with assembled cantilevers tested in a $\pm 1\text{G}$ field in the temperature range of -55C to $+85\text{C}$ indicate that residual stresses associated with assembly contribute to the temperature sensitivity of the bias frequency. Centrifuge testing has been conducted over the range of 0-120G with an indicated non-linearity of less than 2mG.

INTRODUCTION

Accelerometers using quartz tuning fork or beam technology have been under development for several years [1]. Crystalline quartz has been chosen for use in accelerometers because of its known high stability, ease of piezoelectric excitation, and ease of fabrication. The device described here was designed to measure acceleration in the range of $\pm 130\text{G}$ ($1\text{G} = 9.8\text{m/s}^2$) with an overall accuracy of about $\pm 0.5\text{G}$. However, because of the high stability of quartz oscillators, the potential for much higher accuracy exists. Stability better than $100\mu\text{G}$ would allow application as a navigation accelerometer with precision comparable to those currently used in high accuracy navigation devices. Because of the ease of fabrication of the active quartz elements and the relative simplicity of an open loop device such as this accelerometer, the potential cost-performance ratio is attractive.

The design of this sensor was determined primarily by the required operating environmental conditions and to a lesser extent by the desired accuracy and range of the sensor. The ruggedness specification of the sensor was a survival shock amplitude of 1500G, a minimum beam resonance greater than 2KHz, and an operating temperature range of -55C to $+85\text{C}$. The environmental requirements contribute greatly to the difficulty of maintaining accurate calibration of the sensor over long periods of time. As a consequence of these constraints, the sensor was designed to have a lower sensitivity (scale factor) than similar sensors. The ability to accurately correct the sensor output for temperature variation is directly affected by this lower sensitivity. Therefore, the preponderance of the work reported here was dictated by the need to determine the temperature behavior of the sensor and ultimately minimize the influence of temperature on sensor output.

*This work performed at Sandia National Laboratories supported by the U.S. Department of Energy under Contract No. DE-AC04-76DP00789.

SENSOR DESCRIPTION

The quartz digital accelerometer utilizes double-ended, tuning-fork force sensors patented by EerNisse [2] and a cantilever arrangement of two of these tuning forks patented by Koehler [3].

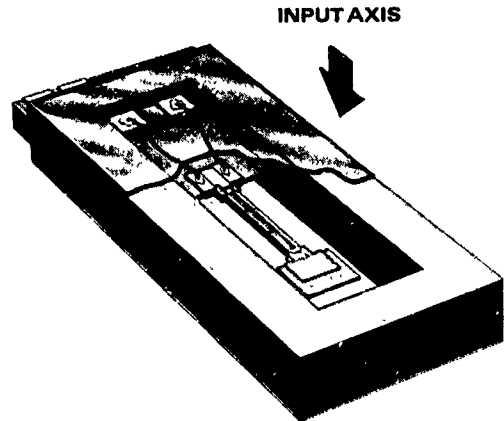


Fig. 1. Quartz crystal acceleration sensor mounted in a vacuum-sealed, alumina package with transparent sapphire windows.

Figure 1 is an illustration of this device. The accelerometer was fabricated from two double-ended tuning forks [4] bonded to beryllium-copper spacers at each end using a non-conductive polyimide adhesive. The tuning forks, which were fabricated from Z-cut quartz, resonate in the fundamental flexure mode at approximately 70KHz and 71KHz, respectively. The electrode pattern and a similar fork configuration have been discussed previously [4]. The cantilever assembly was polyimide bonded to a shelf in the alumina package. Wire leads were attached using thermocompression bonds to the electrode areas of the tuning forks and to bonding pads which feed through the alumina package walls. The package was hermetically sealed in vacuum with sapphire covers. The sealing surfaces were metallized with chromium-gold and the seal was made with tin-gold eutectic solder. The purpose of the transparent sapphire covers is to allow laser tuning of the forks, and possibly the accelerometer proof mass, to obtain manufacturing consistency. The overall dimensions of the completed sensor package is approximately 17mm long by 7mm wide by 3mm high.

The device parameters for the quartz accelerometer are given in Table I.

Table I. Device parameters for the quartz accelerometer.

Proof mass	M = 5.0 mg
Cantilever length	L = 6.8 mm
Tine length	$L_s = 3.7$ mm
Fork thickness	$t_s = 0.15$ mm
Fork tine width	w = 0.18 mm
Spacer thickness	D = 0.13 mm
Quartz density	$\rho = 2650$ kg/m ³
Elastic modulus	E = 7.8×10^{10} Pa

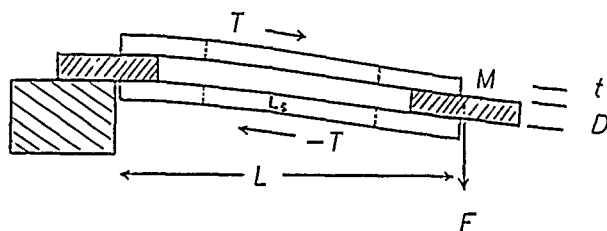


Fig. 2. Schematic representation of quartz sensor operation under an acceleration load.

Operation

Figure 2 shows a cross sectional view of the assembled cantilever configuration containing two double-ended tuning forks. If one assumes fixed joints at the bonded areas between the spacers and tuning forks, a simplified frame (bent) [5] analysis yields the following equation for the tension T in the tines of the upper tuning fork:

$$T = L \cdot M \cdot A / [4(D + t)] \quad (1)$$

where L is the length of the beam to the center of the proof mass, D is the thickness of the spacer, and t is the thickness of the double ended tuning forks. The mass of the proof mass is M and A is the acceleration to be measured.

The frequency of the tuning fork as a function of the tension in the tines is given by [6,7]

$$f = f_0 \sqrt{1 + BT} \quad (2)$$

where

$$f_0 = (a_0 \cdot w / L_s^2) \sqrt{E/\rho} \quad (3)$$

and

$$B = a_s L_s^2 / (E \cdot t \cdot w^3) \quad (4)$$

and where $a_s = 0.294$ and $a_0 = 1.03$ for frequency expressed in Hz.

Using the values for the accelerometer parameters from Table I, the $B \cdot T$ product in (2) does not reach a value of 0.01 until the acceleration is 650G. At this point the error introduced by approximating (2) by the first term in a binomial expansion is approximately 0.25%. Consequently, we have included this approximation in modeling the accelerometer behavior. Thus

$$f \approx f_0 (1 + B \cdot T/2). \quad (5)$$

The scale factor S of the accelerometer is defined as the change in frequency per unit acceleration and is

$$S = f_0 \cdot B \cdot T/A \quad (6)$$

where we have assumed that the frequency changes of the two forks are equal and in opposite directions. These values give a scale factor for the accelerometer of 1.0Hz/G, which is in good agreement with experimental results for the accelerometers which have been built.

Electronic circuits

Figure 3 shows a block diagram of the circuitry associated with the accelerometer. A Pierce oscillator circuit was used for much of the data gathered in these experiments. The oscillator outputs were fed into a mixer circuit which consisted of an exclusive-OR and a filter network. This produced an output

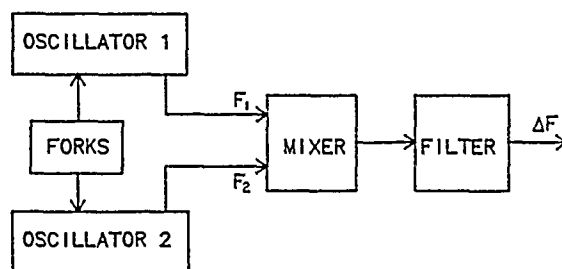


Fig. 3. Accelerometer electronics block diagram.

signal with a frequency equal to the difference between the two oscillator frequencies.

The individual fork frequencies are not important in determining the general behavior of the accelerometer; however, the difference between the individual fork frequencies determines the bias frequency (f_b) of the accelerometer at zero acceleration. The individual forks were usually chosen to yield a bias frequency of approximately 1000Hz. Smaller values for the bias frequency were tried; however, below about 200Hz the two oscillators tended to lock together.

The operational output frequency (Δf) of the accelerometer is

$$\Delta f = f_b (1 + S \cdot A). \quad (7)$$

Characterization of the accelerometer behavior has consisted of determining the effect of various environments on the bias frequency f_b and scale factor S .

Up to this time no effort has been made to control accurately the scale factor or bias frequency except by selection of fork pairs to give the approximate bias frequency desired. However, removing excess gold electrode material from the tine centers by laser tuning after the cantilever beam has been constructed should allow an accurate final determination of the output frequency.

Temperature dependence

There are two major advantages of the dual tuning-fork design. One advantage is that the frequency shifts of each fork add, which increases the sensitivity of the device. The second advantage is that the individual temperature sensitivities of the frequencies of the two tuning forks tend to cancel. Potentially, this cancellation should allow the construction of an accelerometer with a smaller temperature coefficient of bias frequency compared to a design utilizing a single tuning fork. The challenge is to construct a real accelerometer which shows the predicted small temperature sensitivity.

One can calculate the temperature dependence of the individual fork frequency for this particular Z-cut by using the known temperature dependences of the elastic constants and thermal expansion coefficients [8,9] and the expressions for frequency given in (2)-(4). The result of this well known [10,11] calculation is shown in Fig. 4. Although the temperature dependences of the material properties used in the calculations are all monotonic over the range of interest, the resultant calculated frequency temperature dependence is parabolic. This parabolic behavior may be modeled by

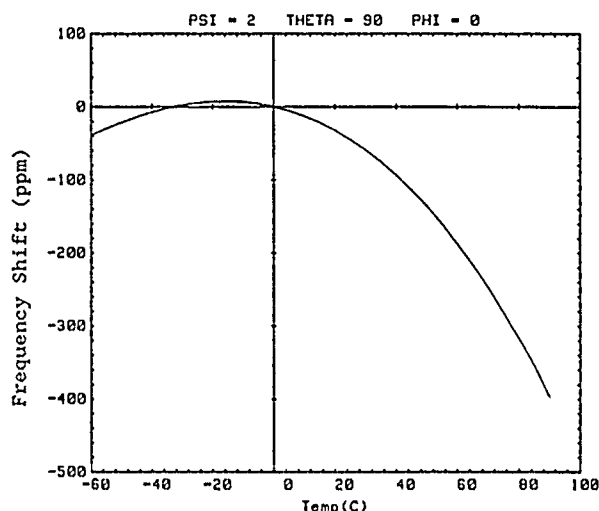


Fig. 4. Calculated variation of frequency with temperature for Z-cut quartz tuning forks.

$$(f_1 - f_{10})/f_{10} = k(T - T_{10})^2 \quad (8)$$

where f_1 is the frequency of the fork as a function of temperature, f_{10} is the frequency of the fork at the turnover temperature T_{10} , and k is a constant.

The effect of temperature on the accelerometer bias frequency and scale factor is estimated by making this same calculation for both double ended tuning forks, assuming they differ by only their tine lengths. Subtracting the results of (5) evaluated for each individual double-ended tuning fork determines the sensitivity of bias frequency to temperature. The scale factor sensitivity to temperature is calculated by applying the same technique using (6). The results of these calculations are shown in Fig. 5.

Complete cancellation of the frequency-temperature dependence does not occur because of the intentional small difference in tine length between

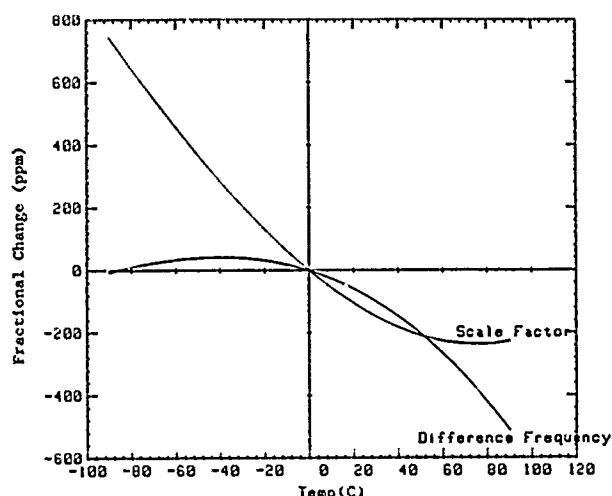


Fig. 5. Calculated sensitivity and bias frequency for the double-ended tuning fork quartz acceleration sensor.

the two forks, which produces the different fundamental frequencies and the non-zero bias frequency. This calculation points out the limit of automatic temperature compensation which one might expect with this design.

EXPERIMENTAL RESULTS

Slew Test

Prior to assembly, impedance analyzer measurements were used to determine the electrical parameters of the individual forks as a function of temperature in a dry nitrogen environmental chamber. These measurements were repeated after the forks were assembled into a cantilever configuration and mounted and sealed in the vacuum package shown in Fig. 1. The measurements were made with the fork axes oriented vertically. This orientation assured that the frequency measured was not significantly affected by the earth's gravitational field.

It should be noted that the individual fork electrical parameters were also dependent on atmospheric pressure. Typically, in dry nitrogen the motional resistance R_m was 200KΩ, the inductance was 7KH, and the Q was 7000, and under vacuum the motional resistance was 50KΩ, the inductance 7KH, and the Q 40000. The resonant frequency increased by about 50Hz when going from atmospheric pressure to vacuum.

Figure 6 shows the measured temperature behavior for two individual double-ended tuning forks before assembly into the cantilever beam configuration (QDA1226). Since frequency difference is important in the accelerometer rather than the fractional changes expressed in (8), the results are plotted as frequency difference $(f_1 - f_{10})$ versus temperature. It is apparent that the behavior is quite similar for the two forks. The turnover temperatures are 16.7C and 18.1C and the parabolic constants k are $-4.12 \times 10^{-8} \text{C}^{-2}$ and $-4.10 \times 10^{-8} \text{C}^{-2}$, respectively.

When these forks were combined into a cantilever and mounted in the vacuum package as shown in Fig. 1, the fork frequency-temperature behavior changed. These results are shown in Fig. 7. In this case the turnover temperatures are 14.8C and 5.3C and the coefficients are $-3.55 \times 10^{-8} \text{C}^{-2}$ and $-3.37 \times 10^{-8} \text{C}^{-2}$.

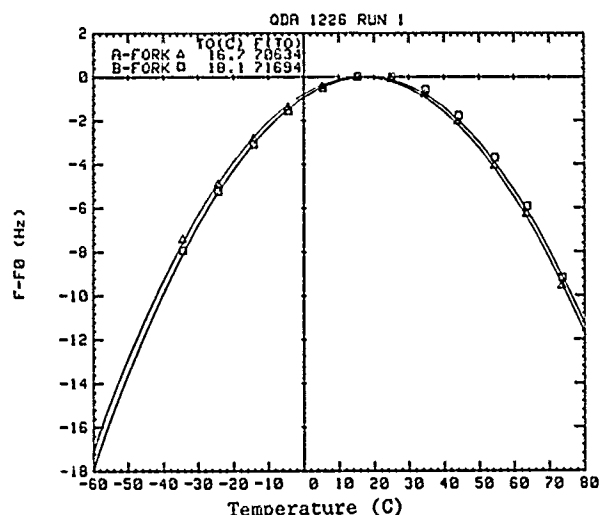


Fig. 6. Measured resonant frequency versus temperature for each of two individual double-ended tuning forks.

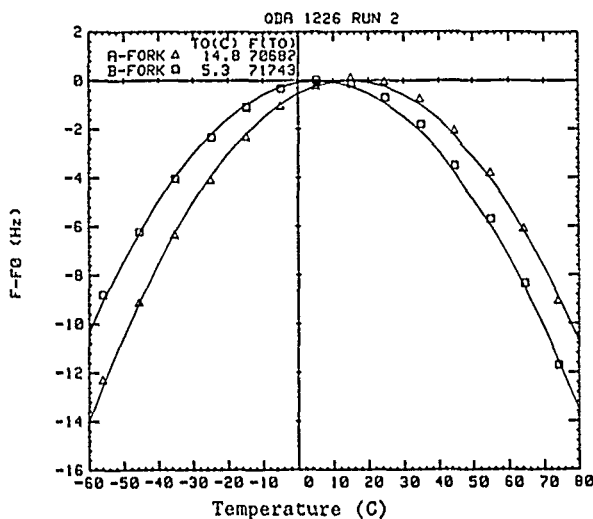


Fig. 7. Measured resonant frequency versus temperature for each fork in an assembled accelerometer (QDA1226).

The difference frequency of the assembled accelerometer was found by subtracting the two curves in Fig. 7 and is shown plotted vs. temperature in Fig. 8. The predicted difference frequency obtained similarly from the frequency curves of the individual forks prior to assembly (Fig. 6) is also plotted in Fig. 8 for comparison. The desired cancellation of the frequency-temperature dependence did not occur primarily because of the apparent shift in the turn-over temperature of primarily one of the forks. This effect is not understood. One potential cause is residual stress in the forks from the assembly process (possibly caused by misalignment of the forks or by mismatch in thermal coefficients of expansion of the adhesive, quartz, and spacers).

Figure 9 shows similar resonant frequency measurements made on another assembled accelerometer (QDA1213). It can be seen that the two forks behave virtually identically with temperature. The resultant

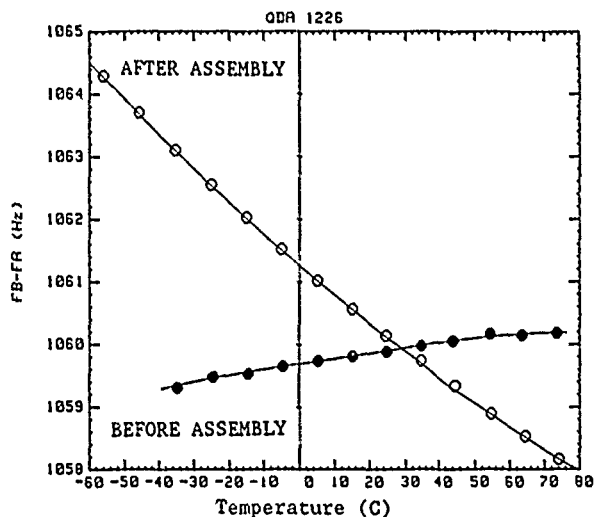


Fig. 8. Difference frequencies for QDA1226 determined by impedance analyzer measurements on the individual forks before assembly (solid symbols) and after assembly into the cantilever beam configuration (open symbols).

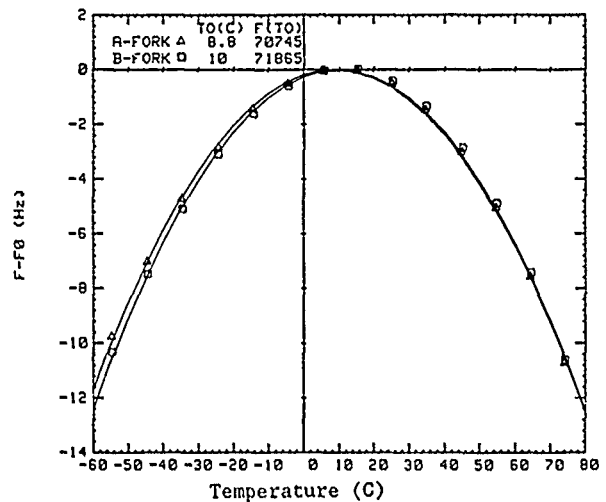


Fig. 9. Measured resonant frequency versus temperature for each fork in an assembled accelerometer (QDA1213).

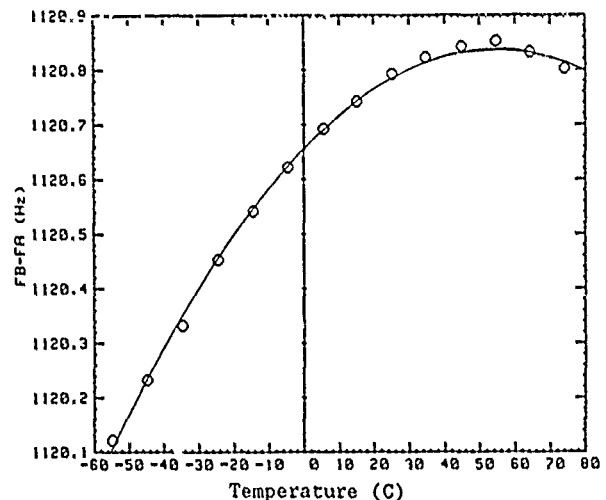


Fig. 10. Output difference frequency of an assembled acceleration sensor (QDA1213) determined by subtracting the individual measured frequencies from Fig. 9.

difference frequency is shown plotted as a function of temperature in Fig. 10. In this case the frequency variation over the complete temperature range is 0.7 Hz compared with the individual fork variations over temperature of about 12 Hz.

Tumble Test

Another measurement which was typically made is called a tumble test. This consisted of generating the output difference frequency of the accelerometer (using the electronic circuits shown in Fig. 3, and a standard frequency counter) while rotating the sensitive axis of the accelerometer in the earth's gravitational field. A range of acceleration values between -1G and +1G were then available for determining the scale factor and bias frequency of the assembled accelerometer. Each series of measurements was taken at temperatures between -55C and +70C. In practice the accelerometer was mounted on a test reference surface with its input axis oriented perpendicular to this surface. The acceleration was found from the cosine of the angle of a normal to the test reference surface with

respect to vertical. This value was corrected for misalignment between the accelerometer sensitive axis and the test surface normal.

Figure 11 shows the tumble test results for QDA1213 at 44°C. A linear least squares fit to the frequency-acceleration measurements produces a point standard deviation of 1.2mHz. The standard deviations of the scale factor and bias frequency are 0.4mHz/G and 0.3mHz respectively.

Figure 12 shows the results of determining the bias frequency by this technique as a function of temperature. The total variation in bias frequency is now less than 0.3Hz. The variation of 0.3Hz corresponds to 270ppm referenced to the bias frequency of 1122Hz. Although not typical, this is better than the theoretically predicted value (450ppm) from Fig. 5 and the value calculated from individual slew test measurements (0.7Hz) from Fig. 10. This improvement may be due to compensation by the electronic circuits used or possible fortuitous compensation from residual stress in the beam assembly. The orientation of the parabola is the same as that predicted, i.e. concave downward.

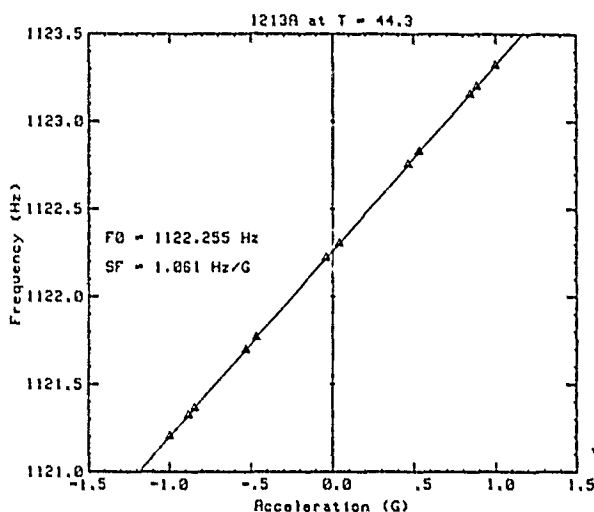


Fig. 11. Tumble test frequency-acceleration data from QDA1213 taken at 44°C.

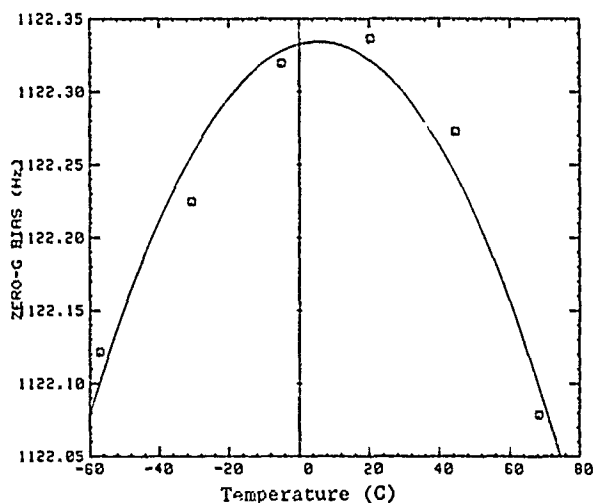


Fig. 12. Zero G bias frequency for QDA1213 determined as a function of temperature from tumble test data least squares line fit. The curve is a least squares parabolic fit of the data.

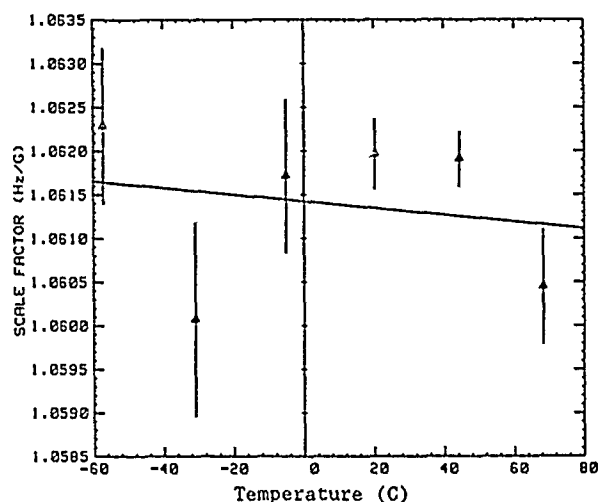


Fig. 13. Scale factor for QDA1213 determined as a function of temperature from tumble test data least squares line fit. The line is a least squares straight line fit of the data.

Figure 13 shows the scale factor determined from the same tumble test data over temperature. The error bars represent one standard deviation determined from the least squares fit to the data. The total variation of the scale factor over the measured temperature range is about 4000ppm compared to approximately 1000ppm predicted from Fig. 5. However, a large portion of this error is due to the measurement error. The measurement error arises primarily from drift of the bias frequency with time which is typically 1-3mHz for this sensor (QDA1213) over the duration of the tumble test measurements (2 hours).

Centrifuge Tests

Several accelerometers were tested using a high precision centrifuge capable of maintaining a stable acceleration to within 10ppm. The accelerometers were tested from 0 to 120G at ambient temperature. The frequency-acceleration behavior for QDA1126 is shown in Fig. 14. It is necessary to plot the deviations from the straight line fit in order to display the non-linearity. The residuals are shown in Fig. 15. The two curves are for increasing and decreasing acceleration levels. Using the bias frequency temperature coefficient, determined independently from tumble test data to be -33mHz/C, the output frequency was corrected for known variations in temperature which occurred during the time data was being collected. It can be seen from Fig. 15 that there is a systematic parabolic variation in the deviations.

A subsequent correction to the centrifuge acceleration value was made by accounting for the centrifuge arm length change caused by the additional load from increasing acceleration. This correction coefficient is 1.16ppm/G. The deviations are re-plotted in Fig. 16 with this additional correction. The result is that the point standard deviation drops from 1.2mHz (no arm stretch correction) to 0.7mHz after making the arm stretch correction. The systematic parabolic deviation is also no longer present. The final point standard deviation represents 6ppm variation from perfect linearity.

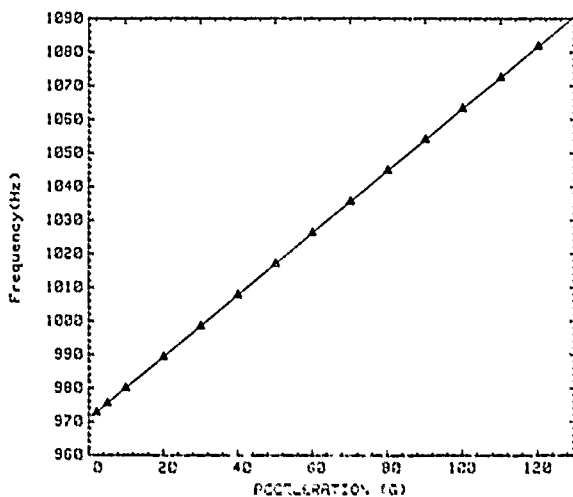


Fig. 14. Centrifuge data taken between 0 and 120G at room temperature for QDA1126.

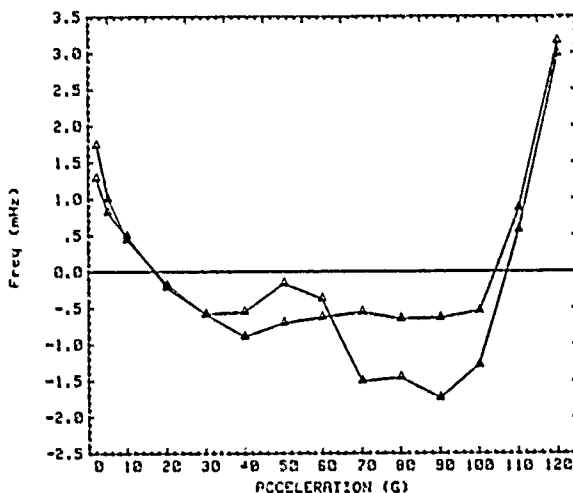


Fig. 15. Residuals from least squares fit to centrifuge data in Fig. 14. The frequency has been temperature corrected but the centrifuge arm length has not been load corrected.

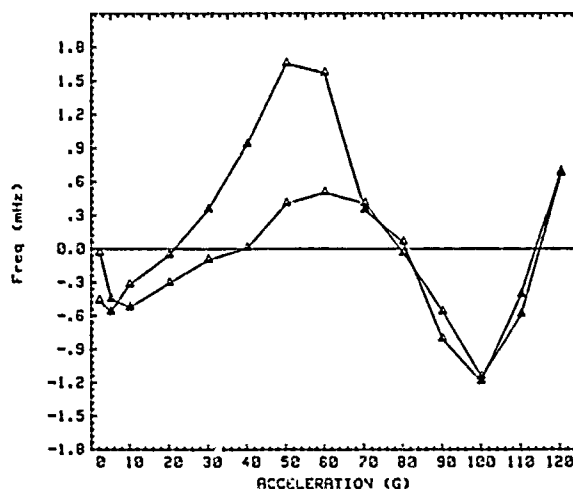


Fig. 16. Residuals from the least squares fit to centrifuge data shown in Fig. 14 but with both temperature corrected data and a 1.16ppm/G centrifuge arm length load correction.

Drop Table Tests

One application of this accelerometer is the measurement of transient acceleration-time profiles with a few milliseconds resolution. To demonstrate feasibility, test accelerometers were mounted on a drop table along with a standard Endevco Model 7701 analog output accelerometer as a reference. Data were collected from the quartz tuning fork accelerometer by measuring the output frequency with a standard electronic counter. The counter was set to measure the period of alternate sensor output cycles with an effective reference clock rate of 1GHz. Since the quartz accelerometer bias frequency was approximately 1KHz the measurement rate using this scheme was about 500Hz.

The drop table tester delivered a shock profile of 30-100ms duration with a maximum amplitude varying from 20G to 80G. Fig. 17 shows the results of a drop which produced a peak of 19.5G and 100ms duration according to the Endevco accelerometer and was measured to be 19.8G and 102ms duration by the quartz tuning fork accelerometer.

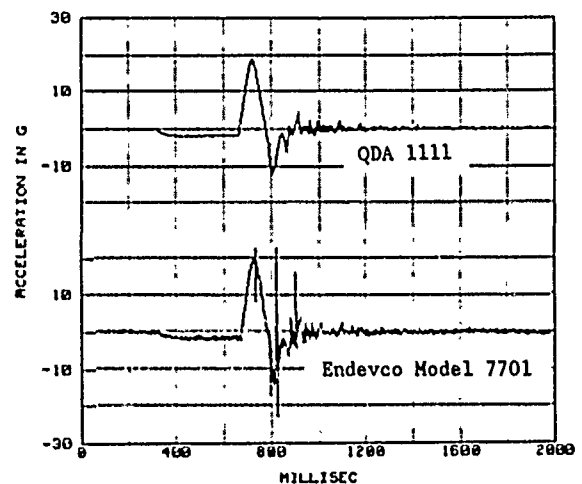


Fig. 17. Comparison of the measured accelerations by an Endevco Model 7701 analog accelerometer and the quartz digital accelerometer in a drop table test.

CONCLUSIONS AND DISCUSSION

The intention of this program is to construct a small inexpensive stable rugged accelerometer for the purpose of measuring acceleration-time profiles. Table II shows a list of characteristics which we have determined for the quartz accelerometer.

Table II. Quartz accelerometer characteristics.

Range	-120G to +120G
Sensitivity	1.0Hz/G
Bias Frequency	1000Hz
Measurement Time	1-2ms
Accuracy	0.2G (1 σ)
Shock Survival	1500G
Vibration Resonance	2200Hz
Operating Temperature	-55C to +85C
Size	16cm ³

The dynamic range of the accelerometer is determined primarily by the ultimate strength of the quartz or the point at which the frequency change becomes sufficiently non-linear to be unuseable. We feel that the range stated here is a conservative estimate of the practical range of this design. The 1500G shock level probably represents the actual operating range although the output will have departed from linearity at this acceleration level. Some preliminary shock data indicate survival above 1500G which correlates reasonably well with an ultimate strength of quartz of about 100MPa.

The sensitivity of the accelerometer is a design trade-off dependent on the desired shock survivability and the lowest acceptable beam resonant frequency. Increasing the sensitivity decreases the shock survival and lowers the beam resonant frequency.

The measurement time of the quartz accelerometer is limited by the lower of two frequencies: the bias frequency and the lowest mechanical resonant frequency of the system. The bias frequency may be varied relatively easily by a proper choice of individual tuning forks, and sensors have been built with bias frequencies between 250Hz and 2000Hz. Higher bias frequencies will allow higher measurement rates at the expense of resolution for a given reference clock rate. A bias frequency of about 1000Hz seems to be a reasonable compromise between high measurement rate and good resolution for our application.

The lowest system resonant frequency has been measured to be 2200Hz. This frequency corresponds to the fundamental vibration mode for the cantilever beam assembly and agrees quite well with that calculated from a frame analysis of the beam.

The accuracy of the accelerometer is primarily determined by the ability to produce sensors with small temperature coefficients of the bias frequency. The accuracy stated here assumes that one can build sensors whose bias frequency varies by no more than ± 0.15 Hz over the operating temperature range. Another means of achieving high accuracy is to compensate for temperature effects with the electronic circuitry. This technique appears promising because the bias frequency temperature coefficient appears to be stable over time but varies from sensor to sensor. This would require that each sensor be calibrated and corrected individually.

The operating temperature range appears to be substantiated by a large amount of data which has been collected between -55C and +85C. The sensor itself will probably survive more extreme temperatures without damage, however, the present electronic circuits are not so robust.

The eventual size of the complete accelerometer will be dictated by the ability to shrink the electronics. The sensor itself occupies significantly less than 1cm^3 . We are currently in the process of hybridizing the electronic circuitry to reach the goal of 16cm^3 .

ACKNOWLEDGMENT

The authors thank Jon R. Bryan for his assistance in fabrication and testing of the accelerometers.

REFERENCES

1. William C. Albert, "Vibrating Quartz Crystal Beam Accelerometer", ISA 28th International Instrumentation Symposium, Vol. 28, no. 1, 1982, pp 33-44.
2. Errol P. EerNisse, "Miniature Quartz Resonator Force Transducer", U.S. Patent No. 4,215,570, Aug. 5, 1980.
3. Dale R. Koehler, "Double Resonator Cantilever Accelerometer", U.S. Patent No. 4,479,385, Oct. 30, 1984.
4. Shih S. Chuang, "Force Sensor using double-ended Tuning Fork Quartz Crystals", Proceedings of the 37th Symposium on Frequency Control, 1983, pp248-254.
5. Harold I. Laursen, Structural Analysis, 2nd Edition, McGraw Hill, New York, 1978.
6. John W. Rayleigh, The Theory of Sound, Vol. I, Second Edition, Dover Publications, New York, 1945, pp255-305.
7. William C. Albert, "Force Sensing Using Quartz Crystal Flexure Resonators", Proceedings of the 38th Symposium on Frequency Control, 1984, pp233-239.
8. J. C. Brice, "Crystals for Quartz Resonators", Rev. Mod. Phys., Vol. 57, 105-146, 1985.
9. R. Bechmann, A. D. Ballato, and T. J. Lukaszek, "Higher-Order Temperature Coefficients of the Elastic Stiffnesses and Compliances of Alpha-Quartz", Proc. of the IRE, Vol. 50, 1962, pp1812-1822.
10. Mitsuo Nakazawa, Yatsuka Nakamura, and Shigemitsu Miyashita, "Frequency-Temperature Characteristics of Quartz Crystal Flexure Bars and Quartz Crystal Tuning Forks", IEEE Trans. on Sonics and Ultrasonics, Vol. SU-26, 369-371, 1979.
11. Virgil E. Bottom, Introduction to Quartz Crystal Unit Design, Van Nostrand Reinhold, New York, 1982.

AN ECONOMICAL TOUCH PANEL USING SAW ABSORPTION

Robert Adler and Peter J. Desmares

Zenith Electronics Corporation
1000 Milwaukee Avenue
Glenview, Illinois 60025

ABSTRACT

Surface waves are readily absorbed when a soft object such as a fingertip touches the substrate. This familiar observation is put to use in a touch-sensing display panel to provide read-outs of finger location and finger pressure.

Reflective array strips placed along the panel edges surround the entire panel area. A short pulse launched, say, along the top edge from left to right arrives at the bottom left corner having a long rectangular envelope, with each point in time corresponding to a specific vertical path across the panel. A finger touch causes an amplitude dip whose timing and depth indicate location and finger pressure.

Low frequencies (4-10 MHz) make low-cost arrays practical. Signal/noise ratio is very high, so the center of the dip can be located precisely; resolution is determined by the associated digital circuits.

I. Earlier SAW Touch Systems

In 1972, R. Johnson¹ patented a touch screen (fig. 1) which used a set of parallel beams of surface waves the way current infrared systems use parallel beams of light: rows of transducers placed at opposing edges of a glass panel were switched cyclically, one transmitting and one receiving transducer which faced each other being turned on at one time. A fingertip placed on the screen would attenuate transmission on one of the many paths, thus revealing the position of the finger.

Other surface wave systems have been described, one based on waves reflected from the fingertip² and others using a stylus³ which contains a receiving transducer. In all these systems, transit time is used to determine position. They all share the need for transducers which cover the entire length of two screen edges, or all four edges in the case of the 1972 patent.

But that 13 year-old system has a very desirable feature: its signal-to-noise ratio is excellent. There are no problems with spurious

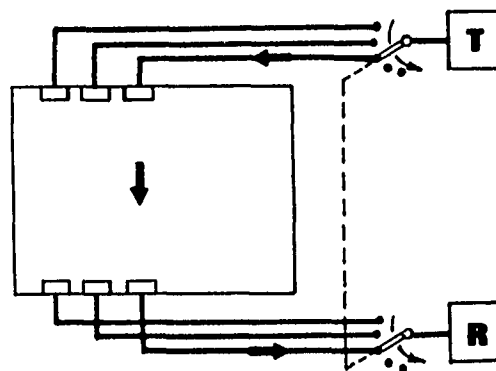


Fig. 1: Johnson's 1972 touch system

reflections or transducer ringing. When the path is clear we get a nice, clean signal, and when that signal goes down there is no doubt that it did.

To retain this attractive feature of an absorption system but avoid the expense of transducers extending the full length of the edges, we decided to borrow from the well-established art of reflective arrays. Fig. 2 shows the principle.

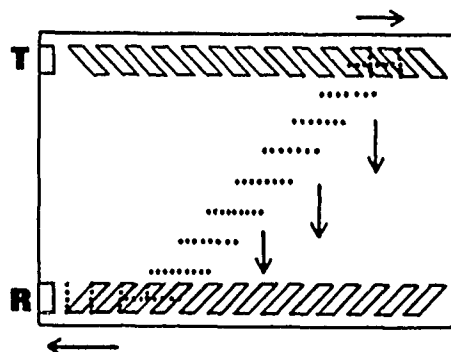


Fig. 2: Touch system using a reflective array

II. The Reflective Array Approach

A transducer in the left top corner of the rectangular panel emits a short burst of surface waves which travel horizontally along the top edge. Within the path of these waves there is an array of partial reflectors positioned at 45° to the direction of wave travel, with their mutual spacing arranged so as to produce cumulative interference. The reflected wavelets traverse the panel vertically. Wavelets coming from reflectors further to the right lag behind those wavelets which were reflected earlier by reflectors further to the left. All wavelets eventually reach a second array along the bottom edge of the panel; this array is an upside-down mirror image of the upper array. It re-directs the incoming wavelets toward the receiving transducer on the left. In the process, those wavelets arriving further to the right, already late, are further delayed by the longer return path.

The signal picked up by the receiving transducer now includes wave portions having experienced widely different transit times. What started as a short pulse is now a continuous signal with a long drawn-out rectangular envelope. Each point on the time axis corresponds to a specific vertical path across the panel. When attenuation is introduced into one such path by a finger touch, a dip appears in the output signal whose timing indicates the horizontal position at which the touch occurred.

Note that this is not a dispersive array. It is designed for a single, fixed frequency. Position is determined strictly by observing transit time.

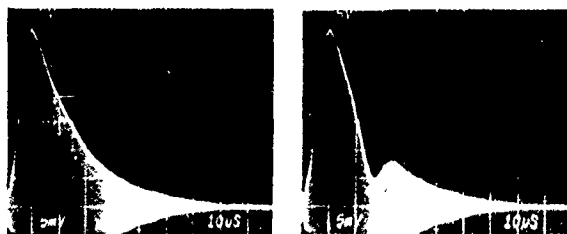


Fig. 3 left: Response of uniform periodic array; right: dip produced by finger touch

Fig. 3 shows on the left the unperturbed output from our first experimental array. A 5 μ sec pulse at 4 MHz was applied to the input transducer. To no one's surprise, the amplitude decreases exponentially with time or distance; the array strips were uniform, so for every inch of horizontal travel a fixed percentage of the remaining power was diverted into a transverse path. A finger touch produced a conspicuous dip (fig. 3 right).

III. Array Weighting

In a practical system, the electronics which process the dip would have an easier job if the unperturbed amplitude were constant, rather than decreasing exponentially with time. To achieve this, we must lower the reflectivity of the array portion near the transducer where most of the input power is still present, but use the highest reflectivity available at the far end to make the best of what little power is left there. If we assume zero dissipation, the power flowing from left to right in the upper array will decrease at a rate equal to the density of the power diverted across the panel. Since we want that density to be constant, the forward power flow p must decrease linearly with distance (straight line in fig. 4). At some point Z , forward power will be completely exhausted. Let us measure distance x along the array backward from point Z ; then the equation of the straight line is simply

$$p = a \cdot x$$

and the power reflectivity R^2 -- the fraction of forward power diverted per unit length of travel -- must increase in inverse proportion to the distance remaining to point Z , or

$$R^2 = 1/ax.$$

This is our design prescription (solid curve).

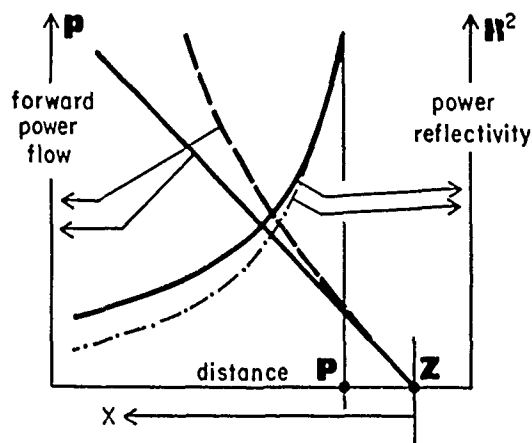


Fig. 4: Design curves for array weighting

Actually, power reflectivity cannot exceed a finite limit, so the real array terminates at some point P short of Z , and whatever forward power remains at P is discarded.



Fig. 5: Finger withdrawal pattern used in 11 x 8" array

In practice we find that there is some dissipation α , a few dB for a 10 inch (25 cm) path. We can correct for this by replacing the straight line $p=ax$ by the function

$$p = (a/\alpha) [(exp \alpha x) - 1] \text{ (dashed curve).}$$

The design prescription then becomes

$$R^2 = (\alpha/a) [(exp \alpha x) - 1]^{-1} \text{ (dash-dot curve).}$$

Analysis shows that the receiving array should be weighted just like the sending array.

To modify the arrays as described, we borrowed again from established weighting techniques. We tried Solie's method⁴ of subdividing the array strips into random dots, we also used random dashes, but generally we employ the method of finger withdrawal familiar in surface wave filters. Fig. 5 shows a typical pattern used in some of our demonstration models, and fig. 6 shows its electrical performance.

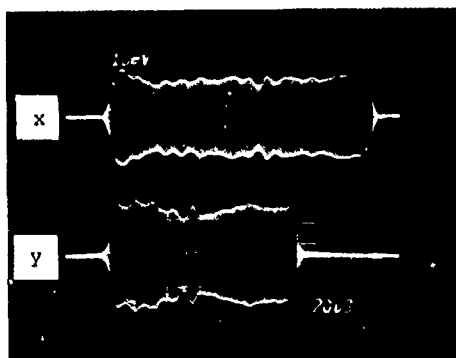


Fig. 6: Unperturbed responses of 11 x 8" array

Note that the envelope is by no means perfectly flat. These imperfections have to do in part with the process by which we make the patterns, and a few words on that subject might be of interest.

IV. Some Practical Details

The surface wave velocity on soda-lime glass is typically 3150 m/sec. At 4.2 MHz, the wavelength is 0.75 mm or 30 mils. To obtain cumulative interference, reflectors positioned at 45° should be spaced by one wavelength λ in the direction of wave propagation. Fig. 7 shows a layout we have used on 14" screens. Note that the perpendicularly-measured center-to-center spacing between reflector strips is $\lambda/\sqrt{2}$ or 0.53 mm (21 mils), hence for a 50% duty factor

the strips should be 0.27 mm or 10.5 mils wide. Their thickness, of course, depends on the material from which they are made, but typically about 5 microns (0.2 mils) are required. We have found it convenient to make the array by screen printing, an inexpensive process widely used to make circuit boards as well as many non-electronic items. The ink we have used is an emulsion of glass frit, a high-density glass powder which crystallizes at about 430°C. We chose this

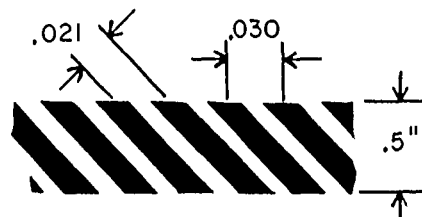


Fig. 7: Detail of reflector strips

because of its proven compatibility with the panel glass, and because our laboratory is familiar with it; no doubt other suitable inks could be found. The essential point is that the positioning accuracy of the screen printing process, probably quite inadequate for making a tiny 500 MHz reflective array compressor, is eminently satisfactory for producing an 11" x 8 1/2" (28 x 22 cm) array of strips spaced by half a millimeter (20 mils).

One attractive feature of any absorption-based system is the pressure dependence of surface wave attenuation: The more strongly we press the finger against the screen, the deeper the dip. In a computer, we might for example use a light touch to run a cursor to a desired position and then take action--replace a character, or switch to a different display--simply by pushing a little harder. It is interesting to remember how the increase in absorption with pressure comes about. The soft tissue of a fingertip has acoustic properties not too different from those of water, with a velocity of about 1500 m/sec for compressional waves. In contact with a solid of much higher impedance, glass, which carries surface waves at more than 3000 m/sec, the surface wave becomes leaky; its power wanders off into the water-like fingertip. The percentage of power so lost depends only on the size and shape of the contact area. Thus it is not the pressure

itself but the area of actual contact between finger and glass that determines attenuation.

Soft materials such as cotton, nylon, rubber or leather gloves work much like bare fingers. Pencil erasers which are usually fairly hard require more pressure. Hard objects such as metal or hard plastic touch the glass in only a few points, regardless of pressure, and do not produce useful absorption. For a given absorber, percentage absorption increases with frequency; the 10 MHz region may be a good choice for small screens. We have gone as high as 13 MHz in experiments, and the increase in touch sensitivity is quite striking.

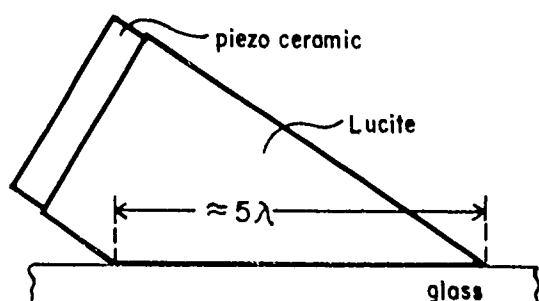


Fig. 8: Cross section of wedge transducer

The transducers we have used (fig. 8) are conventional wedge transducers, made of Lucite wedges with thickness-resonant slabs of PZT-5 or lead metaniobate attached to one side. The bottom of the wedge is later epoxy-bonded to the glass plate. For experiments we use vacuum grease bonds, which are almost as good but permit positioning and re-using the transducer-equipped wedges. The wedges are highly unidirectional and quite insensitive to frequency. The angle between the incident longitudinal wave vector and the glass surface is 33 degrees.

A few words about the electronics, whose task it is to recognize the presence of a dip and determine the time and depth of its lowest point. Typically, because our arrays are still not perfect and in some cases also because there are grease spots or oil spots on the screen, the

response is not quite flat even in the absence of a touch. In a very simple system, this is neglected: One must push hard enough to make the signal drop below a preset threshold. Such a system is cheap but not very sensitive. In the demonstration systems we have built, information on the reference amplitude at many points along the time axis is stored in a digital memory, and current information is then compared with the stored reference.

Signal/noise ratio of the acoustic signal is very high, so the center of the dip can be located quite precisely; resolution is determined by the choice of the associated digital circuits. 320 x 256 resolution, with 16 pressure steps, has been used in demonstration models. A large portion of the electronics is used for both x- and y-coordinates on a time-sharing basis.

All the work reported here was done with non-dispersive arrays, with transit time constituting the variable which corresponds to position. It is also perfectly possible to use dispersive arrays; if this is done, specific frequencies correspond to specific portions of the array, and as a consequence frequency becomes available as a second or alternate means of determining position. Such systems involve many changes with respect to transducers, array precision requirements and, of course, electronics, and all these changes affect the economics of the system; an interesting subject, but one which is outside the scope of this paper.

References

- 1 Ralph Johnson, U.S.Pat. 3,673,327 (1972)
- 2 Alvin Hlady, U.S.Pat. 3,916,099 (1975)
- 3 Paul W. Woo, U.S.Pat. 3,134,099 (1964)
- E. Dieulesaint et al, El. Ltrs. 12, No. 22, p. 586 (1976)
- A. Ishii and S. Hashimoto, 1981 Ultrason. Sympos. Proc., p. 167
- 4 L.P. Solie, 1979 Ultrason. Sympos. Proc., p. 682

RANDOM WALK FREQUENCY FLUCTUATIONS IN SAW OSCILLATORS*

T. E. Parker

Raytheon Research Division
131 Spring St.
Lexington, MA 02173

Summary

In 1983 it was reported that the long term frequency stability of surface acoustic wave (SAW) oscillators is made up of both systematic drift and very long period random frequency fluctuations^{1,2}. A spectral analysis showed that these random frequency fluctuations are random walk in nature and that the magnitude of these fluctuations is greater in delay line oscillators than in resonator oscillators. New data on 25 SAW oscillators is reported. These devices include 19 two-port resonator oscillators in the 400 MHz to 1000 MHz range and six delay line oscillators at 400 MHz. Parameters evaluated for resonators were device power level and frequency, type of transducer metal and preseal bake temperature. For delay lines silicone treated devices and split finger devices were analyzed. It has been confirmed that the SAW devices are the source of the random walk noise and that temperature fluctuations are not the cause of the frequency variations. The level of the random walk noise increases with device frequency and there is a correlation between random walk noise level and flicker noise level for the different device types. However, among "identical" devices the correlation is weaker.

Introduction

In 1983 it was reported that the long term frequency stability of surface acoustic wave (SAW) oscillators is made up of both random and systematic frequency shifts^{1,2}. For well packaged SAW devices the systematic drift is small and comparable in magnitude to the random frequency fluctuations. A spectral analysis of the random frequency fluctuations shows that they are random walk in nature. This means that the spectral density of frequency fluctuations, $S_{\Delta F}(f_m)$, varies

approximately as $1/f_m^2$, where f_m is the modulation or Fourier frequency. This random walk noise exists in both SAW delay line and SAW resonator oscillators, but is generally larger in magnitude in delay line oscillators. The random walk noise is present in both types of oscillators from $f_m \approx 10^{-2}$ Hz to below 10^{-8} Hz, and flicker noise ($S_{\Delta F}(f_m) \propto 1/f_m$) is present from $f_m \approx 10^{-2}$ Hz to $f_m \approx 10^3$ Hz. For delay line oscillators it has been confirmed that the random walk frequency fluctuations are too large to be caused by temperature variations and in one case it has been demonstrated that the SAW delay line itself is the source of the random walk noise¹. Most of the data up till now has been on delay line oscillators.

In this paper we report new data on the random walk noise levels of 25 SAW oscillators. These devices include 19 two-port resonator oscillators in the 400 MHz to 1000 MHz range and six delay line oscillators at 400 MHz. All of these devices were fabricated on 36.7 degree rotated Y-cut quartz, and were packaged in cold-weld TO-8 enclosures which were sealed under high vacuum. The spectral density of frequency fluctuations was obtained from a Fourier analysis of frequency variations recorded as a function of time. One class of such data is the aging curve for the SAW oscillators. This consists of 100 or more data points with each point representing the average frequency for one week. A second class of data consists of 1000 or 2000 frequency values recorded as a function of time with averaging periods ranging from 5 to 200 seconds. The 200 second runs with 1000 points take almost 60 hours and are usually made over weekends. Runs with shorter averaging times are made overnight.

*This work supported in part by U.S. Air Force Contract F19628-84-C00103 with Rome Air Development Center (EEAC).

Among the nineteen resonator oscillators are nine devices at 416 MHz. Six of these were operated with 10 mW of RF power dissipated in the SAW device and three dissipated 1 mW. Of the six high power devices, three resonators were fabricated with copper-doped aluminum transducers³ and three with pure aluminum transducers. The three low power devices also had pure aluminum transducers. Seven 425 MHz resonator oscillators were also evaluated at low power. Three of these were sealed after a low temperature (~260°C for 1 hour) bake rather than the normal high temperature (~350°C for 1 hour followed by 300°C for 15 hours) bake. Of the four normally processed 425 MHz resonators, random walk noise measurements were made on two for both high and low loaded Q (electrical) conditions. All of these devices had pure aluminum transducers. Three 985 MHz resonator oscillators are also being analyzed. These resonators were also fabricated with pure aluminum transducers and are dissipating 1 mW of RF power.

In addition to the nineteen resonator oscillators, data on six new 400 MHz delay line oscillators is also being gathered. Three of these delay lines have split fingers and three have normal quarter wavelength fingers. The three devices with normal fingers were silicone treated⁴.

Experimental Results

416 MHz Resonator Oscillators

Aging experiments on nine 416 MHz resonator oscillators have been completed with over 100 weeks of data on each device. Also at least one weekend run has been made on each oscillator and in some cases two weekend measurements were made. As mentioned above, all of these devices were sealed in cold-weld TO-8 enclosures after the normal long, high temperature bake. Six of the resonators were fabricated with pure aluminum transducers (750 Å thick) and three were fabricated with copper-doped aluminum (0.5 %) transducers of the same thickness. Three of the pure aluminum devices and the three copper-doped aluminum devices were all operated with 10 mW of RF power dissipated in the SAW resonator. This power level is right at the threshold of accelerated aging due to high acoustic strain³. The remaining three pure aluminum devices were operated with 1 mW of RF power dissipated in the SAW resonator.

As expected, the high power devices with pure aluminum transducers showed the greatest systematic frequency drift with shifts ranging from 0.4 PPM to 5 PPM among the three resonators after 100 weeks of operation. The copper-doped aluminum devices showed shifts in the range of 0.3 PPM to 1.4 PPM after 100 weeks of operation. The low power, pure aluminum devices showed the least drift with shifts ranging from 0.3 PPM to 1.1 PPM after 100 weeks of operation.

The results of the spectral analysis of these nine resonator oscillators are summarized in Table 1. The random walk noise is characterized by its magnitude, $S_{\Delta F}(f_m)$, at $f_m = 10^{-6}$ Hz and the slope α , where α is the exponent of the power law dependence on f_m . The values for the magnitude and slope for the random walk noise were obtained from a best fit to the data in the range of $f_m = 10^{-8}$ Hz to 10^{-3} Hz. As can be seen in Table 1 the values of α are close to 2, which is characteristic of random walk noise. The average value of α for all nine oscillators is 2.1. The flicker noise level of each oscillator was also measured and is indicated in the table by the value of $S_{\Delta F}(f_m)$ at $f_m = 1$ Hz. The spectral density curves for two of the oscillators are shown in Figs. 1 and 2.

TABLE 1
SUMMARY OF SPECTRAL ANALYSIS OF 416 MHz OSCILLATORS

Device Number	Random Walk		Flicker $S_{\Delta f}(f_m=1 \text{ Hz})$	
	$S_{\Delta f}(f_m=10^{-6} \text{ Hz})$	Slope		
$P_O = 10 \text{ mW}$ Pure Al	Q1283A	$3 \times 10^6 \text{ Hz}^2/\text{Hz}$	2.2	$3 \times 10^{-4} \text{ Hz}^2/\text{Hz}$
	Q1283D	2×10^7	2.1	1×10^{-4}
	Q1283C	7×10^7	2.0	8×10^{-4}
AVG.		3×10^7		4×10^{-4}
$P_O = 10 \text{ mW}$ Cu - Al	Q1284C	3×10^7	2.0	3×10^{-2}
	Q1284A	1×10^8	1.9	7×10^{-3}
	Q1284D	1×10^7	2.0	6×10^{-4}
AVG.		5×10^7		1×10^{-2}
$P_O = 1 \text{ mW}$ Pure Al	Q1286B	5×10^6	2.2	2×10^{-3}
	Q1283E	3×10^6	2.1	1×10^{-4}
	Q1283F	2×10^7	2.1	3×10^{-4}
AVG.		9×10^6		8×10^{-4}

The data in Table 1 shows that the average random walk noise level was higher in the high power devices than in the low power oscillators. This was not true, however, for

the flicker noise level. Two out of three of the copper-doped aluminum devices showed high levels in both the random walk and flicker noise, but this is not typical of most devices fabricated with copper-doped aluminum transducers. In fact flicker noise data from over 20 other copper-doped resonators shows lower than average noise levels. The copper-doped devices in this study had particularly bad finger definition and therefore are suspected of being not representative of typical devices. The high power devices with pure aluminum transducers show low flicker noise levels yet the average random walk noise level is more than three times higher than the average noise level of the low power devices. This indicates that there may be a power level dependence to the magnitude of the random walk noise, particularly in light of the fact that five of the six devices came from the same quartz substrate. The increased level of random walk noise may be related to the higher systematic frequency drift of the high power devices. A definite conclusion concerning a power dependence will require data from more devices since there is a large scatter in the device to device noise levels.

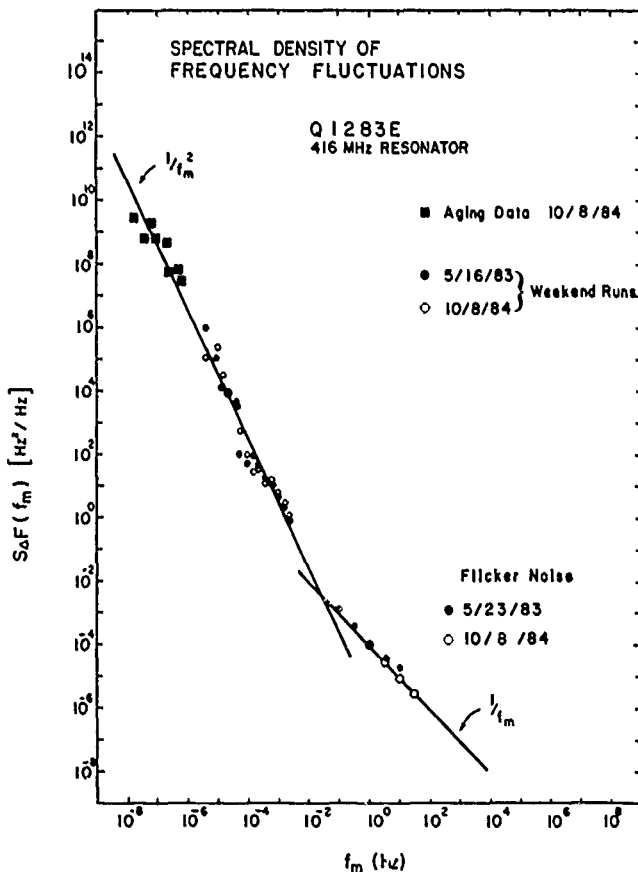


Figure 1. Spectral density of frequency fluctuations for Q1283E (1mW device).

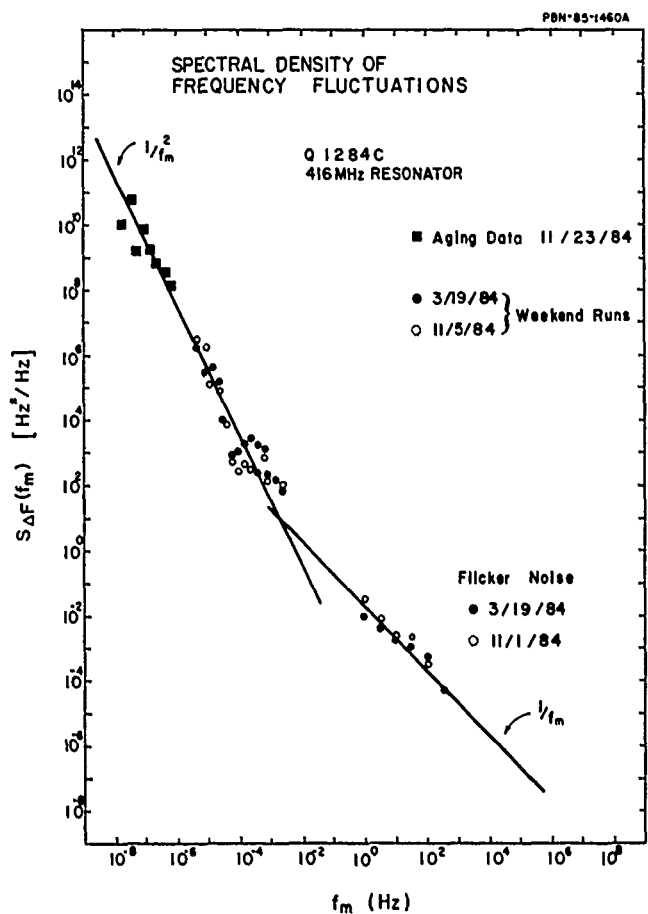


Figure 2. Spectral density of frequency fluctuations for Q1284C (10 mW device with copper-doped aluminum).

One interesting characteristic of the noise spectrum of devices with high random walk noise levels is the presence of significant deviations from the $1/f_m^2$ dependence in certain portions of the Fourier frequency range. Note that for the low power device in Fig. 1, the random walk noise spectrum is reasonably uniform and was reproducible over a period of seventeen months. However, the spectrum for a high power, copper-doped device in Fig. 2 shows significantly more structure in the frequency range of 10^{-4} Hz to 10^{-2} . Also, the structure was not reproducible. This kind of behavior is not unusual for devices with high random walk noise levels.

Random walk ($1/f_m^2$) frequency fluctuations have been reported by Noguchi et al.⁵ in 5 MHz bulk wave quartz crystal oscillators and they were attributed to random walk temperature fluctuations in the ovens which held the oscillators. To determine if this were the case for SAW oscillators, the temperature of each SAW device was recorded

during the aging test. Figure 3 shows a typical plot of temperature fluctuations and fractional change in frequency (with systematic drift removed) as a function of time for one of the 416 MHz oscillators. As can be seen, there is little correlation between the frequency fluctuations and the variations in the temperature. (Temperature data was not recorded before week 30.)

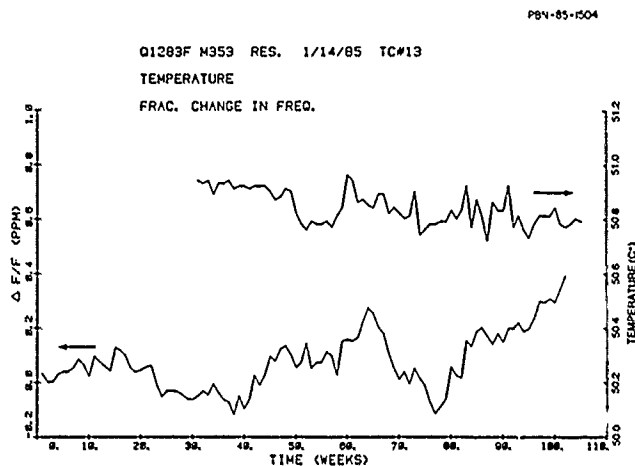


Figure 3. Frequency change and temperature versus time for Q1283F in oven #3.

To be more quantitative, the RMS temperature fluctuations and spectral densities were also calculated from the temperature vs time data of each oscillator. These calculations were useful in determining whether the magnitude of the temperature fluctuations was large enough to influence the oscillator frequencies. To make this estimate, however, it must be taken into account that the recorded frequencies during the aging test are corrected for the observed temperature variations. This is accomplished by making use of the static frequency-temperature characteristic of each oscillator, which was measured before the start of the aging test. This data is used to calculate a second order fit to the frequency vs temperature curve of each oscillator. With this equation, the measured frequency can be corrected for the known temperature variation. If this process could be carried out with absolute accuracy there would be no influence of temperature at all on the corrected aging data. However, the accuracy of the temperature measurement is only $\pm 0.025^{\circ}\text{C}$, and the fit equation may also have some error. To evaluate the influence of temperature on frequency, the RMS fractional frequency fluctuation and RMS temperature fluctuation (for the entire aging period) of each oscillator have been determined and are listed in columns 3 and 1 respectively of

Table 2. Also listed in Table 2 are the first order temperature coefficient of frequency (TCF) for each oscillator (at its average operating temperature), and the estimated fractional frequency variation (in column 4) caused by an effective residual temperature error of $\pm 0.03^{\circ}\text{C}$. The effective residual temperature error includes the combined effect of temperature measurement errors and the inaccuracies in the fit equation. It is used with the first order temperature coefficient (TCF) to calculate the approximate RMS frequency variation (column 4) caused by residual uncorrected temperature fluctuations.

TABLE 2
SUMMARY OF ANALYSIS OF FREQUENCY AND TEMPERATURE
FLUCTUATIONS FOR 416 MHz OSCILLATORS

DEVICE NUMBER	ΔT_{RMS} ($^{\circ}\text{C}$)	TCF (PPM/ $^{\circ}\text{C}$)	MEASURED $\Delta F_{\text{RMS}}/F$ (PPM)	CALCULATED $\Delta F_{\text{RMS}}/F$ (PPM)
OVEN #1	Q1283A	0.10	-0.032	0.084
	Q1283D	0.12	+0.13	0.092
	Q1283C	0.11	+0.13	0.126
	Q1284C	0.21	+0.002	0.083
	Q1284A	0.11	-0.34	0.129
OVEN #3	Q1286B	0.06	+0.60	0.047
	Q1283E	0.06	+0.76	0.027
	Q1283F	0.06	+0.87	0.092
	Q1284D	0.06	+0.06	0.064

As can be seen in Table 2, the actual RMS frequency variations (column 3) for the five oscillators in the oven labeled as #1 are significantly larger than the estimated residual frequency fluctuations (column 4) caused by temperature variations. A spectral density analysis of these temperature variations was also carried out using the same computer software as that used for the frequency variations, and it showed an average slope of $f_m^{-1.8}$, which is fairly close to the typical slope of $f_m^{-2.0}$ for the frequency fluctuations. However, consistent with the RMS data, the magnitude of the power spectral density of the temperature fluctuations is too small to be a source of error for the oscillators in oven #1. The situation in oven #3 is not so clear, however. For example, for device Q1283E the estimated RMS fractional frequency fluctuation is only slightly smaller than the observed fluctuations. The other three devices in oven #3 are not quite so close. The spectral density of temperature fluctuations for oven #3 was also calculated but it shows an average slope of $f_m^{-1.0}$,

which is in substantial disagreement with the $f_m^{-2.0}$ slope of the observed frequency noise. This indicates, that although the magnitude of the temperature effect is large enough to be of concern, the correlation between temperature and frequency variations is still weak. The primary cause for the increased influence of temperature in oven #3 is the larger temperature coefficients of the oscillators. The actual temperature variations in oven #3 were in fact smaller than in oven #1.

To further confirm that temperature fluctuations are not the cause of the random walk frequency fluctuations, weekend measurements were made on a 425 MHz resonator oscillator in which the SAW device could be operated at different ambient temperatures. It was observed that there was no measurable change in the random walk noise level of $S_{\Delta F}(f_m=10^{-6}) = 5 \times 10^6 \text{ Hz}^2/\text{Hz}$ for the same SAW resonator when it was operated near the turnover point and at a temperature 40°C below the turnover point. The static temperature coefficient increased by more than a factor of thirty at the lower operating temperature. If temperature fluctuations were causing the frequency fluctuations, $S_{\Delta F}(f_m)$ would have increased by a factor of 30^2 or 900. The degree of temperature regulation was comparable at both temperatures.

Noguchi⁵ has shown that the dynamic temperature coefficient of an acoustic resonator can also play a role in low frequency frequency fluctuations, so the dynamic temperature coefficient of this device was also measured. It was found to be approximately $15 \text{ PPM}/^\circ\text{C}/\text{SEC}$, which is comparable to what Noguchi observed on the 5 MHz bulk wave resonators. However, even dynamic temperature effects can not explain the level of random walk noise observed in SAW resonators. The levels of $S_{\Delta F}(f_m)$ that Noguchi observed were on the order of 10^9 smaller than that observed on the SAW oscillators, yet the static and dynamic temperature coefficients (fractional change in frequency) of SAW and BAW devices are of the same order of magnitude. If $S_{\Delta F}(f_m)$ for random walk noise shows something comparable to the F_0^4 dependence of flicker noise⁶, where F_0 is the resonator frequency, then it is not surprising that at frequencies well above 5 MHz the inherent noise processes in the acoustic devices begin to dominate over environmentally induced fluctuations.

425 MHz Resonator Oscillators

In addition to the 416 MHz resonators, aging tests and a complete spectral analysis have also been completed on five 425 MHz two-port SAW resonator oscillators. These devices were intended to evaluate the influence of the preseal bake on long term systematic and random frequency fluctuations. The devices were processed in the same fashion as the 416 MHz devices except that three of the resonators received a shorter, and lower temperature preseal bake than the other two. Four of the five devices were fabricated from the same quartz substrate. The short bake consisted of one hour at approximately 260°C just prior to the cold-weld seal. The normal long bake consists of one hour at greater than 350°C followed by 15 hours at 300°C . The devices were sealed after the long soak and dissipated 1 mW of RF power during operation. The aging curves for the five devices are shown in Fig. 4. The curves for the three devices with the low temperature bake are shown in the upper half of the figure and the curves for the two devices with the normal long, high temperature bake are shown in the lower half of the figure. Note the order of magnitude difference in the vertical scales for the two classes of devices. The necessity of a high temperature bake for good long term stability is clearly evident from this data. There is a two order of magnitude difference in total drift between the best and the worst devices. Next to each device identification number in Fig. 4 is the peak temperature seen by that device during the preseal bake. Note how sensitive the magnitude of the systematic

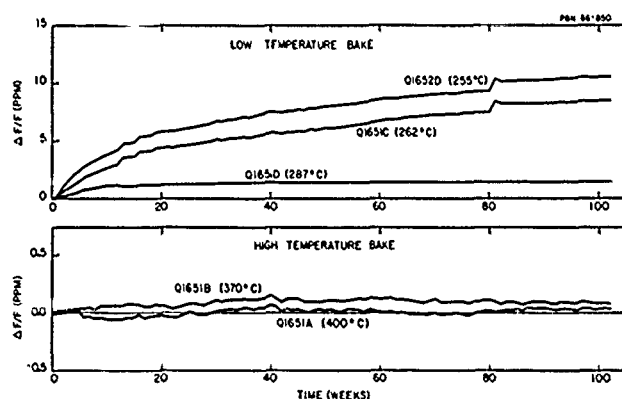


Figure 4. Long term frequency stability of five 425 MHz SAW resonator oscillators.

drift is to the temperature. The frequency drift of device Q1651A has not exceeded 0.1 PPM in over two years of operation. Clearly, a temperature approaching 350°C is required for a long term stability better than 0.5 PPM/year.

The spectral density of frequency fluctuations for device Q1651A is shown in Fig. 5. The $1/f_m^2$ and $1/f_m$ regions are clearly evident and the magnitudes are entirely typical. The data from the two weekend runs, which were made 70 weeks apart, are consistent with each other and with the spectral density data calculated from the aging curve. The spectral density for device Q1652D, the device with the largest systematic drift, is shown in Fig. 6. There are some distinct differences in the nature of this curve as compared to the curve in Fig. 5. The random walk noise region either has a steeper than usual slope ($\alpha = 2.3$) or there are two regions of normal slope ($\alpha = 2.0$) but with different magnitudes. These two possibilities are illustrated by the solid and dashed lines in Fig. 6. The device in Fig. 6 also has a very distorted flicker noise region. All three of the low

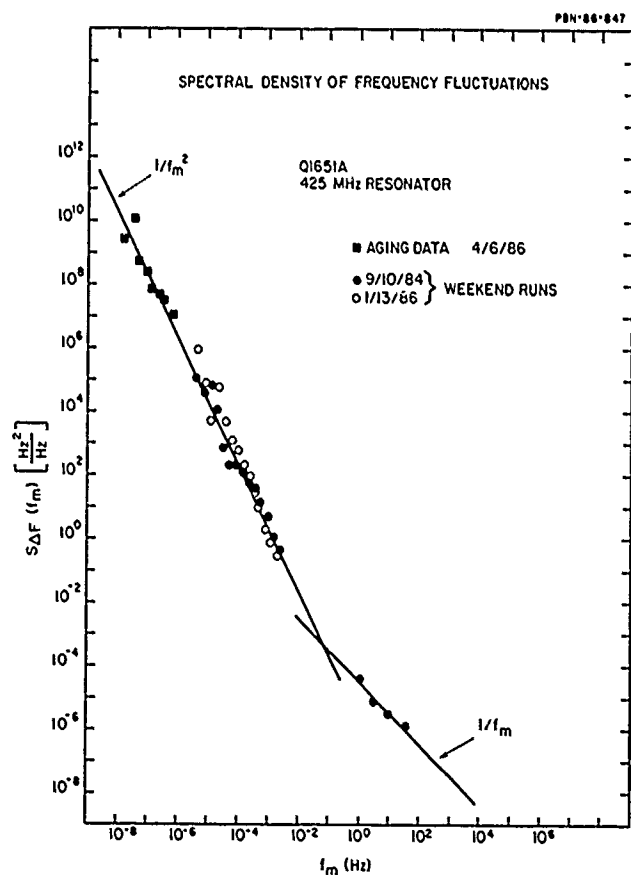


Figure 5. Spectral density of frequency fluctuations for a 425 MHz resonator with a high temperature bake.

temperature bake devices showed some degree of additional steepness in the random walk region indicating a connection between large drift rates and higher than normal random frequency fluctuations. The one hundred fold greater magnitude in systematic frequency drift in Q1652D as compared to Q1651A is accompanied by an approximate one hundred fold greater value for $S_{\Delta F}(f_m = 10^{-6})$. However, this increase in $S_{\Delta F}(f_m)$ results in only a ten fold increase in the RMS frequency fluctuations. (See Eq. 5 which is discussed later.) Thus the systematic frequency drift increases faster than the RMS frequency fluctuations about the drift. Table 3 summarizes the results of the spectral analysis of the five 425 MHz resonators.

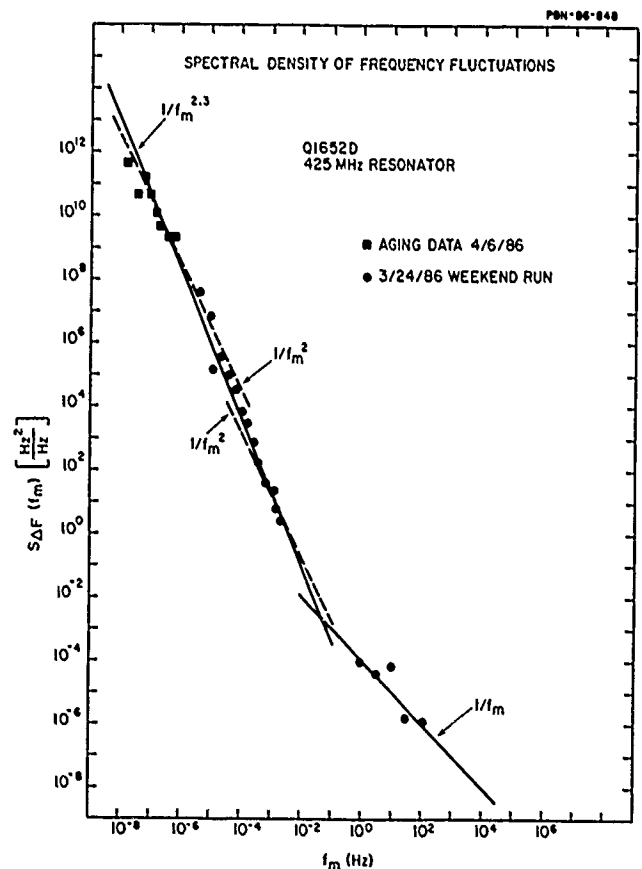


Figure 6. Spectral density of frequency fluctuations for a 425 MHz resonator with a low temperature bake.

TABLE 3
SUMMARY OF SPECTRAL ANALYSIS OF 425MHZ OSCILLATORS

	Device Number	RANDOM WALK $S_{\Delta F}(f_m=10^{-6}\text{Hz})$ Slope		FLICKER $S_{\Delta F}(f_m=1\text{Hz})$
High Temp. Bake	Q1651A	$2 \times 10^6 \text{ Hz}^2/\text{Hz}$	2.0	$3 \times 10^{-5} \text{ Hz}^2/\text{Hz}$
	Q1651B	2×10^6	1.8	4×10^{-5}
Low Temp. Bake	Q1651C	2×10^8	2.3	5×10^{-5}
	Q1651D	5×10^6	2.2	9×10^{-5}
	Q1652D	2×10^8	2.3	1×10^{-4}

Devices Currently in the Aging Test

There are eleven more oscillators that are in the aging test for which a spectral analysis is being performed. The aging period has not been completed yet so the spectral analysis of the aging data has not been performed at this time. However, at least two weekend runs have been completed on each of the eleven oscillators. Five of the eleven SAW devices are two-port resonators and the remaining six devices are delay lines. The two-port resonators consist of two 425 MHz devices, identical to the five resonators just discussed, and three resonators that operate at 984 MHz. All were sealed with the normal high temperature preseat bake in TO-8 enclosures. The six delay lines all operate at 400 MHz and consist of three devices with split-fingers and three with quarter wave length fingers. Except for the finger widths the designs for both type devices are identical. The normal quarter wave finger design is the same one used for the delay lines in references 1 and 2. The quarter wave finger devices were silicone treated⁴ and received a low temperature bake in order to not burn off the silicone polymer. The split-finger devices received the normal high temperature bake.

The two 425 MHz resonators are providing some very interesting results since they have been operated in both a high and low loaded Q, Q_L , condition. Prior to the start of the aging test these two devices were operated in an oscillator circuit in which the loaded Q was equal to approximately $0.55 Q_U$, where Q_U is the unloaded Q. This is the same condition all of the previously discussed resonators were operated in. At least one weekend run and one overnight run were made

on each of the two devices in order to characterize the random walk noise level. The oscillators were then reconfigured for low Q_L operation. This was accomplished by using series matching inductors at each port of the SAW device. Each coil consists of three loops of wire wound on a #30 drill bit. With the matching coils the insertion loss of the resonators was reduced from $\sim 8 \text{ dB}$ to $\sim 4 \text{ dB}$, and the loaded Q was reduced to $\sim 0.3 Q_U$. The reduced insertion loss of the SAW resonators increased the oscillator open loop gain so attenuators had to be added to the circuit to bring the gain back to +3 dB. The attenuation was added after the SAW devices so that the dissipated power remained at 1 mW as was present in the high Q_L configuration. The oscillators were then put into the aging ovens and more weekend runs were performed.

TABLE 4
SUMMARY OF SPECTRAL ANALYSIS OF HIGH AND LOW Q DEVICES

Device Number	RANDOM WALK $S_{\Delta F}(f_m=10^{-6}\text{Hz})$	FLICKER $S_{\Delta F}(f_m=1\text{Hz})$	Q_L
Q1867A	$6 \times 10^6 \text{ Hz}^2/\text{Hz}$	$5 \times 10^{-4} \text{ Hz}^2/\text{Hz}$	8,000
	1×10^7	4×10^{-4}	13,600
Q1867B	3×10^7	1×10^{-3}	6,900
	3×10^7	1×10^{-3}	12,200

Table 4 summarizes the results of the random walk and flicker noise measurements of these two devices for both the high and low loaded Q conditions. As can be seen in the table, there is no significant difference in either the random walk or flicker noise levels between the high and low Q conditions for the two oscillators. Figure 7 shows a plot of the spectral density of frequency fluctuations for one of the devices for both the high and low Q_L configurations. The lack of difference in noise levels is clearly evident. A least mean square fit to the random walk data for both 425 MHz resonators shows that there is less than 1 dB average difference in noise levels for the high and low Q_L configurations. If the random walk or flicker frequency fluctuations were being caused by oscillator components other than the SAW resonator, the low Q_L noise would be approximately 5 dB higher than the high Q_L noise. This can be shown by using the relation between phase shift, $\Delta\phi$, and frequency change, ΔF , in an oscillator

$$\Delta\phi = 2\pi\Delta F\tau \quad (1)$$

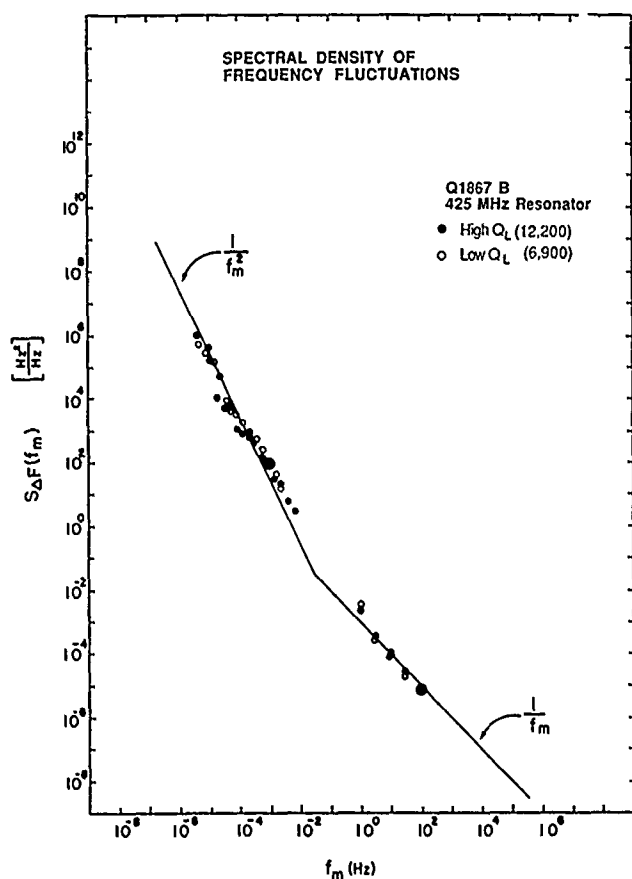


Figure 7. Spectral density of frequency fluctuations of Q1867B in both the high and low Q_L configurations.

and the relation between Q_L and group delay, τ .

$$Q_L = 2\pi F_O \tau / 2 \quad (2)$$

where F_O is the resonator frequency. Equations 1 and 2 yield

$$\Delta F = (F_O / 2Q_L) \Delta \phi \quad (3)$$

In terms of spectral densities, this becomes

$$S_{\Delta F}(f_m) = (F_O / 2Q_L)^2 S_{\phi}(f_m) \quad (4)$$

where $S_{\phi}(f_m)$ is the spectral density of phase fluctuations of the oscillator amplifier or other non-SAW components. The Q_L^{-2} dependence in Eq 4 means that by reducing the loaded Q by a factor of 1.8, a 5 dB increase in noise will occur if the dominant source of random walk or flicker noise is from phase fluctuations in the electronic components of the oscillator. The data in Table 4 and Fig.

7 therefore demonstrates that the SAW devices are the source of both the random walk and flicker noise. The only possible exception would be the highly unlikely possibility that the phase fluctuations through the electronic components were so sensitive to the matching circuits at the SAW devices that the external phase fluctuations would actually decrease by 5 dB under the low Q conditions. For flicker noise, this possibility has already been ruled out through other work at Raytheon Research Division.

Preliminary random walk data for the three 984 MHz oscillators has also been obtained. Figure 8 shows the spectral density data for one of these devices. The random walk noise level is higher in the 984 MHz devices than in the 400 MHz resonators, but it is difficult to be more specific at this time. Like the low temperature bake resonators, the three 984 MHz devices all show a transition to a region of noise with a slope steeper than $1/f_m^2$. However, these devices did receive a normal high temperature bake. Clearly, something different is happening in

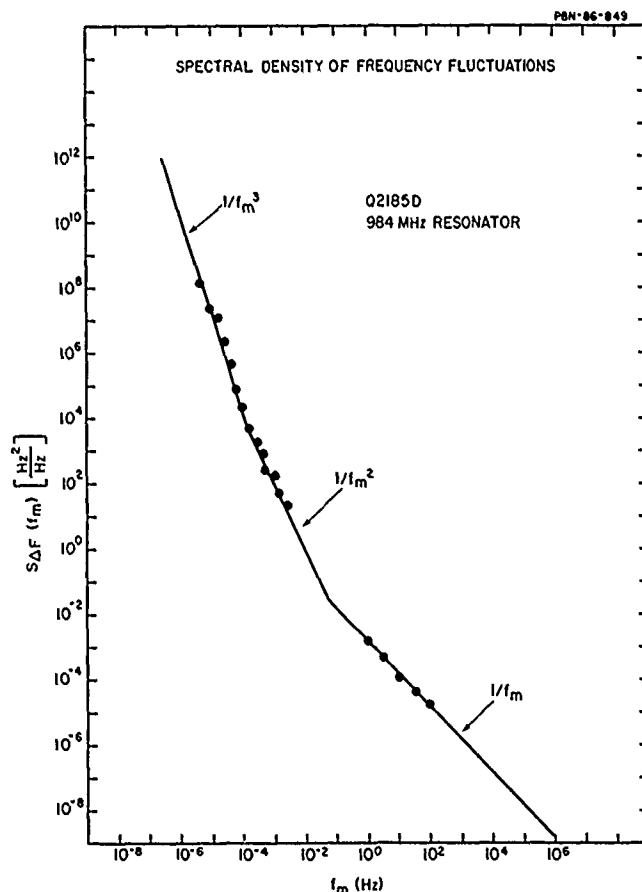


Figure 8. Spectral density of frequency fluctuations for a 984 MHz SAW resonator oscillator.

these high frequency devices as compared to the 416 and 425 MHz resonators. A spectral analysis of 45 weeks of aging data suggests that the steep $1/f_m^3$ slope does not continue to lower frequencies and that the values of $S_{\Delta F}(f_m=1 \times 10^{-6} \text{ Hz})$ are in the range of 10^8 to $10^9 \text{ Hz}^2/\text{Hz}$. However, a more definitive answer regarding the random walk noise levels of the 984 MHz resonators will have to wait for the conclusion of the aging test when the results of a more complete spectral analysis will be available.

The preliminary data for the three new silicone treated delay lines and the three new split-finger devices in the aging test are included in Fig. 9 which is discussed in the following section.

Correlation Between Random Walk and Flicker Noise

Figure 9 is a plot of $S_{\Delta F}(f_m=1 \times 10^{-6} \text{ Hz})$ versus $S_{\Delta F}(f_m=1 \text{ Hz})$ for all of the normally processed resonators discussed in this report. The vertical axis represents the level of random walk noise, and the horizontal axis represents the level of flicker noise. Also included in Fig. 9 are the data from the six new delay line oscillators in the aging test, and data from Reference 2 for a large number of untreated 400 MHz delay line oscillators and four additional silicone treated 400 MHz delay lines. The data in Fig. 9 clearly shows that there is a strong degree of correlation between random walk noise and flicker noise. Among the various types of devices,

fabrication variables, and operating conditions that are represented by the data in Fig. 9 there is a clear correlation between devices that give low flicker noise levels and the presence of low random walk noise levels. However, among "identical" devices in Fig. 9 the correlation is not so strong, and the large amount of scatter which is typical for noise processes is present for both random walk and flicker noise.

Clearly, there are design or fabrication parameters in SAW devices that influence both the random walk and flicker noise levels at something approaching a dB for dB correlation rate. Resonators are generally quieter than delay lines for both types of noise, but Q or group delay can not be totally responsible for this. The silicone treated delay lines have the same delay time as the untreated devices, yet they have lower noise levels. The copper-doped resonators have a much larger phase slope (group delay) than the delay lines, yet two of the three have noise levels comparable to delay lines. The split-finger delay lines have normal flicker noise levels but slightly higher than average random walk noise levels. However, among devices of the same design and fabrication or operation variables, there is less correlation between random walk and flicker noise levels. Clearly, there are other factors which can influence the random walk and flicker noise levels independently.

Calculation of Standard Deviations

One of the most important reasons for measuring the random walk noise level of oscillators is that you can then calculate the standard deviation (RMS frequency variation) of the oscillator frequency for a given set of measurement conditions. For example, over a two year period both systematic and random frequency fluctuations may play an important role in the overall frequency stability of a SAW oscillator. In many cases now the systematic drift is small compared to the random frequency fluctuations, so the long term stability is dominated by the effects of the random walk noise. In such a case it would be very useful to be able to predict the average amount the frequency will vary about the systematic drift rate. Random walk noise measurements made on the 416 and 425 MHz resonators have shown that the random walk noise level is generally constant over a two year aging test. The new resonators in the aging test will provide additional data on

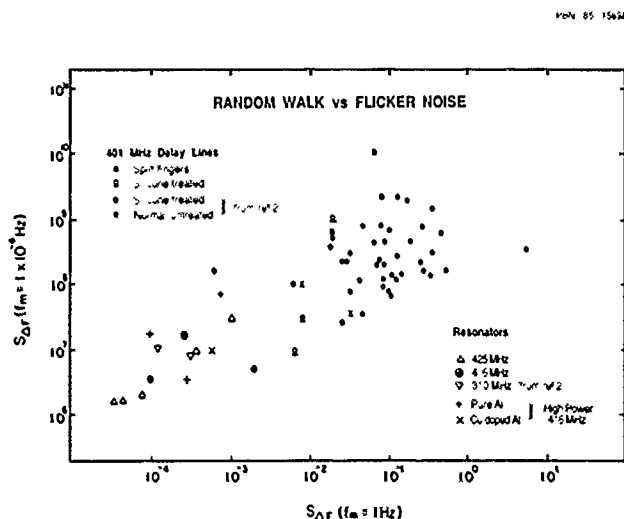


Figure 9. Correlation of random walk and flicker noise in SAW oscillators.

how representative the random walk noise levels measured very early in the aging test are as compared to the levels throughout the entire aging test. A very useful tool would be the ability to measure the random walk noise in a weekend test and then predict the approximate level of RMS frequency fluctuations over a period of two or more years.

If the spectral density of frequency fluctuations, $S_{\Delta F}(f_m)$, varies exactly as $1/f_m^2$ the standard deviation, $\Delta F_{RMS}/F_0$, can be calculated by using the equation below,

$$\Delta F_{RMS}^2/F_0^2 = (2\pi)^2 K_2 \tau N / 12 F_0^2 \quad (5)$$

where

$$S_{\Delta F}(f_m) = K_2/f_m^2$$

Here F_0 is the oscillator frequency in Hz, τ is the measurement period (counter gate time) in seconds, and N is the number of measurements.

If the exponent of f_m in the spectral density is not exactly 2 there is no closed form solution for the standard deviation and it must be calculated with a computer. Such a program is available and it has been used to make these calculations⁷. Specifically, the standard deviation has been calculated for $N = 100$, $\tau = 604800$ seconds (one week) and for the exponent of f_m varying from 1 to 2.8. The spectral density of frequency fluctuations is expressed as

$$S_{\Delta F}(f_m) = K_\alpha / f_m^\alpha$$

where α can vary from 1 to 2.8. Equation 5 applies when $\alpha = 2$. Figure 10 shows the values for the coefficient C as a function of α . The coefficient C is used in the equation below to calculate the RMS frequency deviation for given values of $S_{\Delta F}(f_m=10^{-6})$ and α .

$$\Delta F_{RMS}/F_0 = C[S_{\Delta F}(f_m=10^{-6})]^{1/2}/F_0 \quad (6)$$

The standard deviation is given here in terms of $S_{\Delta F}(f_m)$ at $f_m = 10^{-6}$ Hz rather than K_α as a matter of convenience since the random walk noise levels have been given as $S_{\Delta F}(f_m=10^{-6}$ Hz) elsewhere in this report. K_α can be calculated from $S_{\Delta F}(f_m=10^{-6}$ Hz) by using the relation

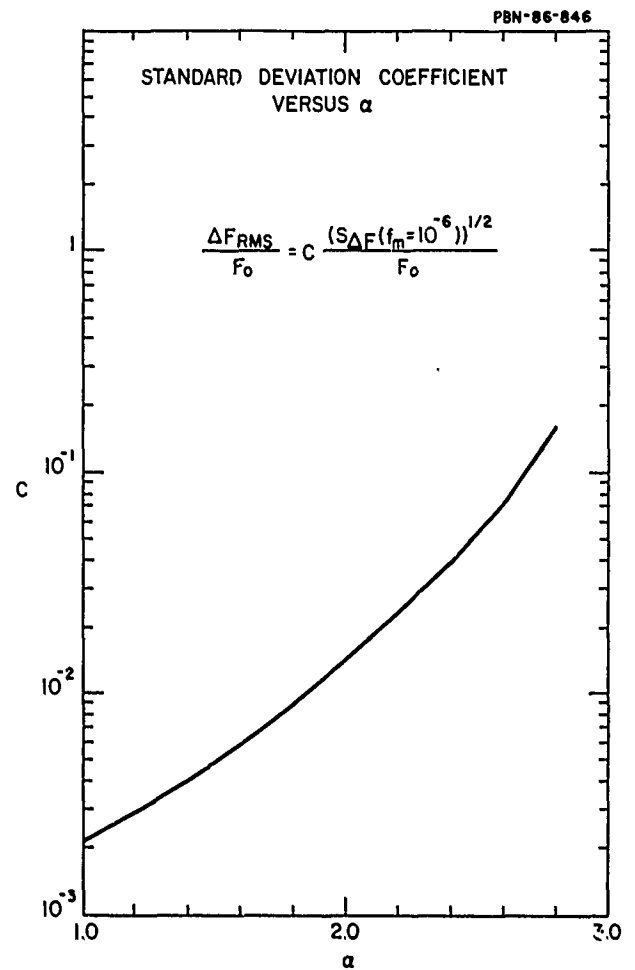


Figure 10. Standard deviation coefficient as a function of α for $\tau=604800$ seconds and $N=100$

$$K_\alpha = S_{\Delta F}(f_m=10^{-6}) / (10^{-6})^\alpha \quad (7)$$

Note that for $\alpha = 2$ and by converting $S_{\Delta F}(f_m=10^{-6}$ Hz) to K_2 , the same results are obtained from Fig. 10 as from Eq. 5.

The values of $N = 100$ and $\tau = 604800$ seconds were chosen for the calculations since they represent the conditions for the aging measurement. A one week gate time is not actually used, but the readings for each day are averaged over a week. This is equivalent to a one week gate time. The results in Fig. 10 show that the calculated RMS frequency variation for a given value of $S_{\Delta F}(f_m=10^{-6}$ Hz) is fairly sensitive to the value of α . If α varies from 1.8 to 2.2 the RMS frequency variation will increase by slightly more than a factor of 2.5. As was stated earlier, the spectral density of

frequency fluctuations on SAW oscillators does not always decrease exactly as $1/f^2$, and therefore some knowledge of the exact slope must also be available if accuracies better than a factor of four are to be obtained in calculating $\Delta F_{RMS}/F_0$.

The standard deviation was calculated from the spectral density data for the fourteen resonator oscillators for which the aging test has been completed and this was compared to the actual standard deviation as determined from the aging data. The calculated standard deviation averaged about a factor of two higher than the measured value. A possible cause for this discrepancy is that a least mean square fit to the time domain data was used to remove systematic drift. This probably artificially reduced the measured standard deviation.

Conclusions

The spectral analysis of aging data on the resonators discussed in this report has provided new information on low frequency noise processes in SAW resonator oscillators. It has demonstrated that the noise in resonators is random walk in nature and that the SAW resonator is the source of the frequency fluctuations. Temperature fluctuations have been ruled out as a cause for this noise in all but the most quiet devices. Fabrication and operation parameters can influence these noise levels and it has been observed that parameters that lead to high systematic drift rates can also increase the level of random frequency fluctuations. The random walk noise level appears to increase with resonator frequency and this increase may be on the order of F_0^4 , but more data needs to be gathered before a definite quantitative relationship can be identified. A definite correlation between random walk and flicker noise levels has also been observed. Device parameters that give low flicker noise levels also tend to give reduced random walk noise levels. However, among "identical" devices there is considerable device to device scatter in the noise levels and the correlation between random walk and flicker noise is less evident.

The cause or causes of both random walk and flicker noise are very poorly understood at this time and no definite conclusions can be drawn. However, the connection between high systematic drift rates and increased levels of random frequency fluctuations may

be significant. Random fluctuations in relaxation processes or migration of material (aluminum in the transducers ?) on the active surface could certainly lead to random frequency fluctuations. In order to understand the causes of random frequency fluctuations the characteristics of these noise processes must first be observed and the device parameters that influence the noise levels must be identified.

Acknowledgements

The author would like to acknowledge the valuable assistance of Don Lee in the writing and using of some of the computer programs. The valuable assistance of Joe Callerame, Gary Montress, Mert Bennett, and John Lang are also gratefully acknowledged.

REFERENCES

1. T. E. Parker, "Random and Systematic Contributions to Long Term Frequency Stability in SAW Oscillator," Proc. of the 1983 IEEE Ultrasonics Symposium, p. 257 (1983).
2. T.E. Parker, "Very Long Period Random Frequency Fluctuations in SAW Oscillators," Proc. of 37th Annual Frequency Control Symposium, p. 410 (1983).
3. W. R. Schreve, R. C. Bray, S. Elliot, and Y.C. Chu, "Power Dependence of Aging in SAW Resonators," Proc. of 1981 IEEE Ultrasonics Symposium, p. 94 (1981).
4. T. E. Parker, "1/f Phase Noise in Quartz Delay-Lines and Resonators," Proc. of the 1979 IEEE Ultrasonics Symposium, p. 878 (1979).
5. Y. Noguchi, Y. Teramachi, and T. Musha, "1/f Frequency Fluctuations of a Quartz Crystal Oscillator and Temperature Fluctuations," Proc. of the 35th Annual Frequency Control Symposium, p. 484 (1981).
6. T. E. Parker, "1/f Frequency Fluctuations in Acoustic and Other Stable Oscillators," Proc. of 39th Annual Frequency Control Symposium, p. 97 (1985).
7. D. L. Lee (private communication)

REAL TIME, INTERACTIVE SAW FILTER COMPUTER
AIDED DESIGN AND ANALYSIS IMPLEMENTATION

S.M. RICHIE, C.D. BISHOP, and D.C. MALOCHA, University
of Central Florida, Department of Electrical
Engineering and Communication Sciences, Orlando, FL
32816.

ABSTRACT

Design automation and integration with device fabrication is a major element towards the goal of increasing a SAW fabrication facilities' productivity. Towards this end, a high level interactive design and analysis tool for both bidirectional and three-phase unidirectional filters has been developed at the University of Central Florida, called UCF-SAWCAD. This system provides the design engineer with a graphical interface, menu driven operation, and access to all of the system's many analysis modules. The present system runs on a DEC VAX 11/750 and contains over 10,000 lines of FORTRAN code. UCF-SAWCAD's system capabilities include many different FIR design techniques and the analysis models include calculation of a frequency dependent lumped parameter model used in conjunction with superposition analysis. A 10 node electrical network may be included with the transducer analysis for modeling external circuits and parasitic effects. This paper will present UCF-SAWCAD's capabilities and future CAD goals. A design example with experimental verification is included.

I. INTRODUCTION

The integrated circuit industry has invested considerable effort in computer aided design and analysis systems (CAD). The benefits of such systems are increased capability and productivity. These effects are achieved through greatly reduced design time which provides fast multiple design iterations. This allows for investigation into design-performance parameter relationships which, conversely, increase design capability. Productivity is also increased since analysis is performed before design commitment. As each technology becomes adequately understood and satisfactory design models developed, the CAD systems become more highly automated. The CAD goal is to achieve complete design automation. Such a system must integrate design and analysis from the device description level to the device fabrication level. In the integrated circuit industry such efforts are leading to a "silicon compiler". The UCF-SAWCAD system is a similar effort for SAW filters. This system is presently evolving towards a "SAW compiler" called UCF-SAWCOM. To facilitate the evolution to such a system, an overall CAD philosophy of vertical and lateral integration capability has been adopted. Vertical integration capabilities allow a hierarchical approach to system architecture while lateral integration capabilities allows design depth, and the easy integration of new devices. Some elements of such a system architecture implementation are standard nomenclature, standard menu layout, data consistency between different data modes, and minimization of intermediate data.

II. A. UCF-SAWCAD OVERVIEW

The objective of this work is to provide a real-time, interactive, graphics oriented design synthesis and analysis CAD system for SAW transducers and filters. Such a system has the advantage that many design alternatives may be investigated in a short time. Rapid iterations provides insight into design-performance parameter relationships. Such insight

increases design productivity by reducing design time and allowing analysis prior to fabrication commitment. The interactive nature of the system, implemented with menu-driven operation, provides quick and easy parameter alterations and redesign. In order to achieve these objectives, the computer model used must, necessarily, have quick execution to be real-time. Also, modular architecture of the system software is used to achieve the menu-driven interactive operation. UCF-SAWCAD's graphical display allows for the examination of filter performance over broadbands, thus allowing quick assimilation of data. The system is presently about 10,000 lines of FORTRAN 77 code. It may be installed on any DEC VAX running under VMS with a TEKTRONIX 401X compatible graphics.

II. B. UCF-SAWCAD CAPABILITIES

UCF-SAWCAD is capable of designing and analyzing two-phase bidirectional filters and three-phase unidirectional filters. The present system execution is controlled via five different menus. These are the "main menu", "FIR design" menu, "transducer analysis" menu, system "utilities" menu, and the "graphics" menu. Figure 1 shows the control menu hierarchy.

UCF-SAWCAD ORGANIZATION

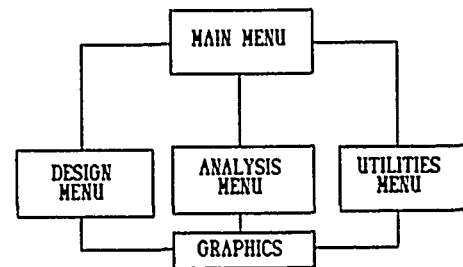


Figure 1 - Menu Hierarchy

The UCF-SAWCAD main menu allows the user to begin a design session, begin an analysis session, or use the system utilities. Figure 2 shows the main menu.

```

UCF-COM-88093      UCF-SAWCAD Version 2.5      14137  26-MAY-86
-----
LAST INPUT FILES NONE      SYSTEM INFORMATION      LAST OUTPUT FILES NONE
-----
      <<  MAIN  MENU  >>
      (FIR design synthesis)  (analyze transducer)
      (utilities)            (quit UCF-SAWCAD)
      COMMAND >
  
```

Figure 2 - UCF-SAWCAD Main Menu

The "FIR synthesis" menu options are shown in Figure 3. These design techniques include Eigen function design for single transducer weighted filters [1], Eigen function design for near optimal simultaneous design of both filter transducers [2], Remez Exchange algorithm design of single weighted transducer filters [3], Window function design [4], and withdrawal weighting design capabilities. The resultant data from an analysis session is a tap weight versus time/space design file. The "Transducer analysis" menu options are shown in Figure 4.

```

UCF-SAWCAD Version 2.5      15145 26-MAY-86
=====
1 LAST INPUT FILE: NONE      <<< SYSTEM INFORMATION >>>
1 LAST OUTPUT FILE: NONE
=====
<<< TRANSDUCER DESIGN SYNTHESIS MENU >>>
(E)Eigen synthesis: both transducers      (M)Eigen synthesis: single transducer
(R)Remez-exchange: both transducers      (I)Remez-exchange: single transducer
(D)Window functions      (G)Graphics      (O)Return to MAIN MENU
COMMAND : ==>

```

Figure 3 - UCF-SAWCAD FIR Synthesis Menu

```

UCF-SAWCAD Version 2.5      15145 26-MAY-86
=====
1 LAST INPUT FILE: c323.v12      <<< SYSTEM INFORMATION >>>
1 LAST OUTPUT FILE: c323.vr2
1 DATA TYPE: Frequency      NUMBER OF DATA SAMPLES: 4096
=====
<<< ANALYSIS MENU >>>
(I)Tap factor extraction      (B)Triple transit analysis
(M)Matching network calculation      (I)Initialize/reset network elements
(C)Complete frequency response analysis      (T)Transducer model parameters
(V)View network      (G)Graphics menu      (O)Return to MAIN MENU
COMMAND : ==>

```

Figure 4 - UCF-SAWCAD Transducer Analysis Menu.

The modeling of the transducer is separated into two processes which link the electrical excitation of the electrodes with the electrode pattern effects to predict the resulting surface wave propagation. The first segment of the model uses the electrode pattern and the electrode charge distribution to identify the frequency response of the transducer using the theory of superposition [5]. The second segment of the modeling identifies a lumped parameter equivalent circuit model of the entire transducer for prediction of the electrical interactions of the transducer and the source and load. Figure 5 shows the equivalent circuit used for a bidirectional transducer or for one port of a 3-phase unidirectional transducer.

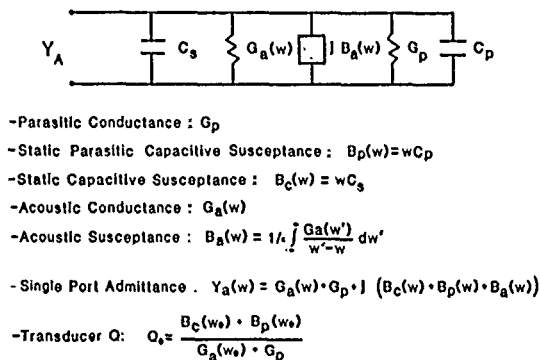


Figure 5 - Single Port Equivalent Circuit

The lumped parameter model includes the frequency dependent acoustic conductance, the Hilbert transform susceptance which is calculated via a double FFT implementation of the Hilbert convolution integral, and the interelectrode capacitance. The

acoustic conductance is calculated from the tap weight data and the knowledge of the implementation structure. Beam centered or slant apodization structures are analyzed using track analysis. Amplitude weighted structures, which are often good approximations to apodized structures, are analyzed via faster closed form representations. During execution of the "complete response analysis" option, in addition to this single node parallel element model, any R, L, or C may be added in any configuration up to defining an arbitrary 10 node electrical network around the transducer equivalent circuit. During the analysis the equivalent circuit may be graphically displayed at any time via the "view network" option. This capability allows for the accurate modeling of parasitic, package, and material effects. Based on this model UCF-SAWCAD will calculate device parameters including insertion loss, apodization loss, S_{11} , and transducer Q. UCF-SAWCAD will also calculate two possible, independent, two-element reactive matching networks which may be included in the analysis. Alternately, the user can define any other matching network or no matching network at all. For three-phase UDTs, both phasing and matching configurations are calculated by UCF-SAWCAD. After analyzing both transducers the complete filter response can be calculated and graphically displayed via the appropriate selections from the main menu. A simple triple transit model is also included for a complete filter response.

Once the design is accepted, a structure module may be invoked to generate the filter structure based on the design file data. This structure may be graphically view and manually altered by the user. After the structure has been accepted, a MANN 3600 format tape may be generated using the data from the structure module via a separate mask tape generation module.

In support of the design and analysis facilities, a fourth menu is available which provides system utilities. These include Fourier transform, file I/O, and file manipulations.

III. DESIGN EXAMPLE

UCF-SAWCAD is illustrated via the design and analysis of a high performance bidirectional filter. Several figures of the design process are included to illustrate the features and operation of UCF-SAWCAD. The design specifications for the filter are:

-6 dB bandwidth = 6.0 MHz
-40 dB bandwidth = 6.9 MHz
(shape factor = 1.15)
center frequency = 60.0 MHz.

From the main menu of UCF-SAWCAD the analysis menu is entered. The design approach of implementing the filter by dividing equally the processing between the two transducers was pursued due to the low shape factor requirement. Therefore, the eigen synthesis of both transducers option is selected from the menu as shown in Figure 6. Figure 7 shows the verification screen for the design information that has been entered by the user.

The eigen design routine then allows the two transducer design files to be saved. Next the filter design analysis can be initiated by first calculating the transducer equivalent circuit model by entering "T" from the analysis menu. This will require verification of the material parameters as shown in Figure 8.

```

UCF-600-00001      UCF-600-00001 Version 2.5      10/15 20-01-00
=====
1  <<< SYSTEM INFORMATION >>>
1  LAST INPUT FILE: NONE      LAST OUTPUT FILE: NONE
1
1  <<< TRANSDUCER DESIGN SYNTHESIS MENU >>>
1
1  (E)igen_synthesis: both transducers      (N)Eigen_synthesis: single transducer
1  (R)esponse_exchanges: both transducers   (I)response_exchanges: single transducers
1  (D)isplay window functions      (G)raphics      (O)utput return to MAIN MENU
1
1  COMMAND: ==> E

```

Figure 6 - Design Technique Selection

```

/*****
<<< DESIGN PARAMETER VERIFICATION >>>
*****/

```

```

6.0 dB Bandwidth is 6.02 MHz
40.0 dB Bandwidth is 6.90 MHz
Transition Bandwidth is 0.44 MHz
Center Frequency is 60.0 MHz
Sampling Frequency is 240.0 MHz
Passband Slope is -2.5 dB

```

```

(A)ccept Values
(M)odify Values
Enter >

```

Figure 7 - Eigen Design Parameters

```

*****
Material Parameter Definitions
*****
ENTER coupling coefficient      1  [ 0.5000E+01 ] ==
ENTER substrate velocity (m/sec) 1  [ 3925. ] ==
ENTER beam width (to wavelength) 1  [ 75.00 ] == 35
ENTER substrate capacitance (pF) 1  [ 3.700 ] ==
ENTER electrode duty factor      1  [ 0.5000 ] ==
ENTER metal resistivity (ohm/cm) 1  [ 0.0000 ] ==
ENTER input/output impedance (real,imaginary) 1
REAL: 1  [ 50.00 ] ==
IMAG: 1  [ 0.0000E+00 ] ==

OPTIONS: (A)ccept values
          (M)odify values
(a) ==>

```

Figure 8 - Material Parameter Screen

In order to calculate the acoustic conductance and susceptance, the structure to be used to implement the design must be known. The four options for apodized structures are shown in Figure 9. Also, amplitude weighted structures may be analyzed in lieu of apodization resulting in decreased accuracy and decreased execution time.

```

Choose an apodization structure

(1) Beam Center Overlap
(2) Generalized Distance Weighting
(3) Beam Spread Overlap
(4) Beam Aligned Overlap

CHOICE ==> 1

```

```

Enter the apodization slope (0 to 1).

SLOPE ==> .95

```

Figure 9 - Structure Choice Screen

Analysis of apodized structures is performed via a track analysis which requires longer execution times as the number of tracks is increased. For 4096 point resolution with 25 tracks, the calculation takes 4 minutes and 18 seconds. The user may alter resolution or the number of tracks in order to trade off execution times versus accuracy. Due to the specifics of the track analysis, the accuracy saturates beyond 25 tracks for this design. During analysis the acoustic admittance may be viewed. Figure 10 shows the calculated acoustic conductance with the experimental transducer conductance overlayed. Figure 11 shows the calculated acoustic susceptance with the transducer susceptance overlayed.

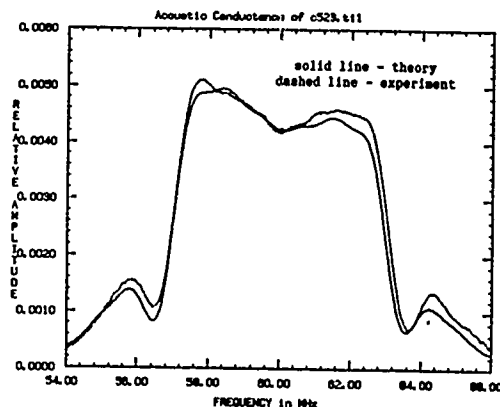


Figure 10 - Conductance Plots

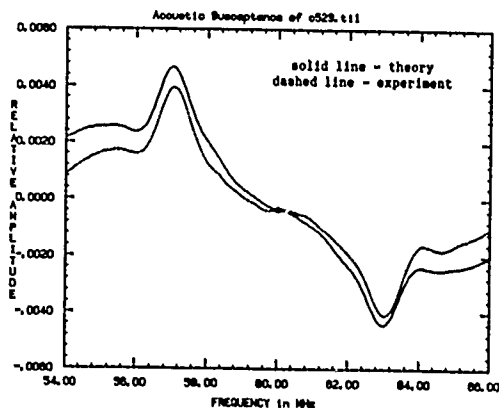


Figure 11 - Susceptance Plots

Once the transducer equivalent circuit has been calculated a data screen is displayed. As shown in Figure 12 among the data presented to the user is the effective number of electrode pairs (vector tap sum), electrode capacitance, thin film loss, input impedance, mismatch loss, transducer Q, apodization loss and total transducer loss.

```

*****
<<< Transducer Model Design Information >>>
*****

Tap sum = 16.87 Vector tap sum = 10.55 Sum of tap weights squared = 9.24

Acoustic conductance = 4232.643 uhoas.
Electrode capacitance = 30.436 pF.
Thin film conductance = 67.061 microhms.
Thin film loss = 0.049 dB.
Thin film resistance = 0.01476 Megohms.

Input admittance conductance = 4.301 mhoas. Susceptance = 11.194 mhoas.
Input impedance resistance = 29.906 ohms. Reactance = -77.642 ohms.
|S11| real = -0.358 Imaginary = 0.426
Amplitude of S11 = 0.721 Mismatch loss = 3.182 dB.
Apodization loss = 0.644 dB
Transducer Q = 2.60 Total bidirectional transducer loss = 6.944 dB

<<< Return to continue >>>

```

Figure 12 - Transducer Model Information

After the transducer model has been calculated, the user can have the system calculate matching networks or calculate the transducer complete response. This analysis is of the unmatched filter. During the complete transducer analysis the electrical network and transducer may be used to calculate the information shown in Figure 13 at any frequency from 0 to f_s .

Once the complete response of the transducer is finished, this data is saved. Similarly, the second transducer is analyzed. Its transducer model information is shown in Figure 14.

The complete filter response including triple transit analysis is performed by selecting triple transit analysis from the analysis menu as shown in Figure 15.


```

1 LAST INPUT FILE: c523.k11  <<< SYSTEM INFORMATION >>>  LAST OUTPUT FILE: NONE
1 DATA TYPE: Frequency  NUMBER OF DATA SAMPLES: 4096
1
1 <<< Electrical Network >>>
1 <<< Design Information >>>
1
1 Frequency = 60.0 MHz
1 Input Impedance: Zin = 29.84 + j77.90
1 S11 = -0.2584 + j0.6261
1 |S11|^2 = 0.3204
1 |S11|^2 = 0.4776
1 Total Insertion Loss = 6.944 dB
1 Apodization Loss = 0.6936 dB
1 Parasitic, Bidirectional & other Dissipative Losses = 3.000 dB
1 Mismatch Loss = 3.191 dB
1
1 Options = 1: Calculate network response
1           2: Recalculate parameters at new frequency
1           3: Modify network
1           4: Quit and return to main menu

```

Figure 13 - Center Frequency Data

```

=====
1 Transducer Model Design Information
=====
1
1 Top sum = 17.12 Vector top sum = 17.26 Sum of top weights squared =
1
1 Acoustic conductance = 4251.540 unms.
1 Electrode capacitance = 30.870 pF
1 Thin film conductance = 68.245 mS/mho.
1 Thin film loss = 7.0487 dB
1 Thin film resistance = 0.01442 Megohms
1
1 Input admittance conductance = 4.720 mms. susceptance = 15.275 mms
1 Input impedance & resistance = 29.820 ohms reactance = -j77.227 oh
1 S11 = real = -0.2584 imaginary = 0.6261
1 Magnitude of S11 = 0.720 mismatch loss = 3.179 dB
1 Apodization loss = 0.626 dB
1 Transducer Qs = 2.61 Total bidirectional transducer loss = 6.074 dB
1
1 <<< return to continue >>>

```

Figure 14 - Second Transducer Model Data

```

=====
1 UCF-SAWCAD Version 2.5 15:41 24-MAY-84
=====
1 <<< SYSTEM INFORMATION >>>
1 LAST INPUT FILE: c523.k12  LAST OUTPUT FILE: c523.f02
1 DATA TYPE: Frequency  NUMBER OF DATA SAMPLES: 4096
1
1 <<< ANALYSIS MENU >>>
1
1 Array factor extraction 1: Triple transit analysis
1 Matching network calculation 2: Initialize/reset network elements
1 Complete frequency response analysis 3: Transducer model parameters
1 View network 4: Graphics menu 5: Return to MAIN MENU
1
1 COMMAND: => 3

```

Figure 15 - Complete Filter Response Analysis

This analysis requires the knowledge of the time delay between the two transducers' centers. Also any unaccounted losses may be included at this time. Figure 16 shows the triple transit modeling and the entry of a value of 3.44 dB of unmodeled losses. The complete filter response data may be saved.

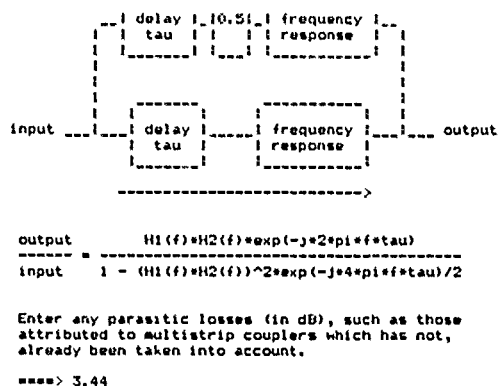


Figure 16 - Triple Transit Model Screen

IV. EXPERIMENTAL RESULTS

The filter was implemented with a slanted apodized structure and fabricated on 128° Y-cut X-propagating LiNbO₃. The transducers are separated by a multistrip coupler whose effects are not accounted for in the model. Figure 17 shows the narrowband filter response theoretical response with the experimental response data overlaid (dashed line). There is good correlation between the experimental data and theory. The triple transit level is accurately predicted.

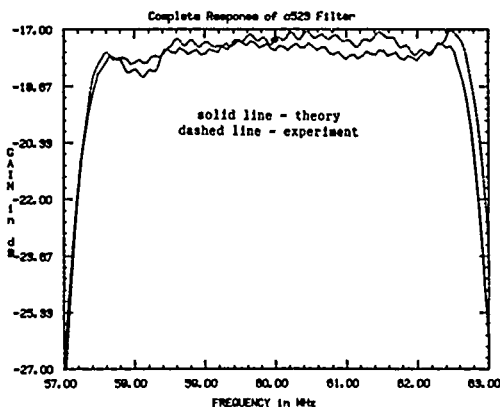


Figure 17 - Narrowband Filter Response

The broadband filter response is shown in Figure 18. The sidelobe level of the experimental response is limited by rf feedthrough and spurious acoustic energy. There is good correlation of theory and experiment over the null bandwidth.

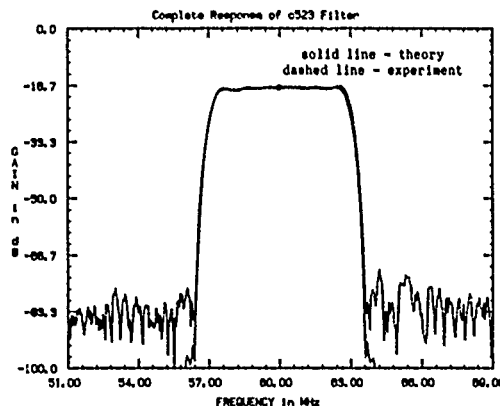


Figure 18 - Wideband Filter Response

Table 1 shows comparisons of theory versus experiment for the filter design parameters.

Parameter	UCF-SAWCAD	Experiment	%Error
Ins. Loss	13.9 dB	17.3 dB	20
6 dB width	5.85 MHz	5.85 MHz	0
40 dB width	6.90 MHz	6.88 MHz	.3
$G_{in}(f_0)$	4.29 mms	4.78 mms	10
$B_{in}(f_0)$	11.2 mms	12.0 mms	7
passband slope	0.65 dB	0.60 dB	8

Table 1

V. DISCUSSION AND CONCLUSIONS

This paper has presented the capabilities of a SAW computer aided design system for design and analysis of SAW filters. When coupled to structure layout program modules and mask generation software, which are available to UCF-SAWCAD, a SAW filter can be designed through mask layout in a few hours with confidence of meeting design specifications.

An example of a low shape factor filter was theoretically and experimentally compared and shown to have good agreement. The device had its largest parameter error in insertion loss. Although unverified, the additional loss may be due to an inefficient multistrip coupler used in the filter. This seems reasonable since the conductance and susceptance of both transducers is predicted well.

UCF-SAWCAD's data format is general and, therefore not being specific to SAW devices, allows for the analysis modules to be used for general signal processing design and analysis. Specifically, the time-frequency domain analysis can be used for arbitrary signal or system analysis or applications.

Currently, efforts are continuing to move the entire SAWCAD system to an IBM PC XT or AT. This will provide a more economical system and produce a SAWCAD workstation environment. Preliminary investigations indicate the system will have no degradation of real time performance when moving from the current VAX 11/750 system to the IBM PC AT.

Based on our current work, it appears the next generation CAD tools will achieve some level of a SAW compiler. This next generation will require sophisticated decision programs to yield near optimum design with little or no user input.

ACKNOWLEDGEMENTS

The authors would like to thank the members of the Solid State Devices and Systems Laboratory for aid in computer program development. A special thanks to Ben Abbott for his contributions to the transducer models.

REFERENCES

- [1] C.F. Vasile, "A Numerical Fourier Transform Technique and Its Application to Acoustic Surface Wave Bandpass Filter Synthesis and Design", IEEE Trans. on Sonics and Ultrasonics, Vol. SU-21, pp 7-11, Jan. 1974.
- [2] C.D. Bishop and D.C. Malocha, "Non-Iterative Design of SAW Bandpass Filters", 1984 Ultrasonics Symposium Proceedings, pp 18-21, Nov. 1984.
- [3] N. Morimoto, Y. Kobayashi, and M. Hibino, "An Optimum SAW Filter Design Using FIR Design Techniques", IEEE 1980 Ultrasonics Symposium Proceedings, Vol. 1, pp 298-301, Nov. 1980.
- [4] F.J. Harris, "On The Use of Windows For Harmonic Analysis With the Discrete Fourier Transform", IEEE Proceedings, Vol. 66, pp 51-83, Jan. 1978.
- [5] S. Datta, B. Hunsinger, and D. Malocha, "A Generalized Model for Periodic Transducers with Arbitrary Voltages", IEEE Trans. on Sonics and Ultrasonics, Vol SU-26, pp 235-242, May 1979.

ON A CONTINUOUS REPRESENTATION OF THE MODE SHAPE IN ACOUSTIC SURFACE WAVE RESONATORS

H.F. Tiersten and D.V. Shick
Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, New York 12180-3590

Abstract

The usual representation of the mode shape along the transmission path in acoustic surface wave resonators is determined by cascading a transmission matrix over a large number of reflectors. This representation is extremely cumbersome to use in perturbation integrals that arise in the calculation of the change in resonant frequency due to biasing states. Consequently, a continuous representation of the mode shape along the transmission path is obtained from the solution of the linear difference system of transmission equations for the periodic array. Although the procedure was developed for convenience in evaluating perturbation integrals for surface wave resonators subject to biases, it applies to any cascaded transmission matrix with small impedance discontinuity. Consequently, it may be useful in the design of large reflecting arrays for surface wave resonators.

1. Introduction

In order to perform an accurate calculation of the change in frequency of a surface wave resonator due to a given acceleration field and support configuration or any biasing state for that matter, using an existing perturbation integral¹, the entire surface wave mode shape must be determined. Although the surface wave decay functions with depth and transverse modal behavior are relatively simple to treat in the perturbation integral, the usual representation of the mode shape along the transmission path is extremely cumbersome to use because it is obtained by cascading a transmission matrix over a very large number of reflectors in a typical resonator. However, since the transmission matrix has a very small impedance discontinuity across each reflector in the array, it is possible to obtain an accurate continuous representation of the mode shape along the transmission path for convenient use in the perturbation integral.

In this paper the desired continuous representation of the mode shape along the transmission path is derived from the solution of the linear difference system for the periodic array. The continuous representation is obtained by expanding the solution of the linear difference system for the periodic array in the small decay parameter resulting from the small impedance discontinuity across a reflector and allowing the element number to be a continuous variable. The resulting continuous representation of the surface wave mode shape along the transmission path, which consists of a sum of two exponentials, is extremely accurate because the impedance discontinuity across each reflector is extremely small in surface wave resonators. The continuous representation obtained here is for a particular transmission matrix for an array of reflecting strips because that matrix was available to us from earlier work^{2,3}. However, the procedure for obtaining the continuous representation is applicable to any cascaded transmission matrix with small impedance discontinuity. Although the procedure was developed for convenience in evaluating perturbation integrals for surface wave resonators subject to biases, it may possibly be useful in the design of large reflecting arrays for surface wave resonators.

2. Surface Wave Equations for an Array of Reflecting Strips

A schematic diagram of a cross-section of an array of reflecting strips is shown in Fig.1 along with the numbering scheme. The plated regions are noted by \bar{n} preceding the unplated region n . The incident region is denoted $n=0$ and the exit region is denoted $n=N$ for the case of N strips. In this section we briefly present the approximate one-dimensional surface wave equations and edge conditions in one scalar variable for each type of region shown in Fig.1 for strips of width $2w$, which is large compared to $\ell+d$, for the lowest transverse mode. These equations were essentially derived in Ref.2 utilizing some results from Ref.3, and are simply presented here with relatively brief explanations.

In Refs.2 and 3 the approximate surface wave equations are obtained by writing

$$u_j = \alpha_j(x_2) \hat{\psi}(x_1, x_3, t), \quad \varphi = \alpha_4(x_2) \hat{\psi}(x_1, x_3, t), \quad (2.1)$$

for an unplated region and

$$\bar{u}_j = \bar{\alpha}_j(x_2) \bar{\psi}(x_1, x_3, t), \quad \bar{\varphi} = \bar{\alpha}_4(x_2) \bar{\psi}(x_1, x_3, t), \quad (2.2)$$

for a plated region, where

$$(\alpha_j, \alpha_4) = \sum_{m=1}^4 C^m(A_j^{(m)}, B^m) e^{i\beta_n \xi x_2}, \quad (2.3)$$

from the known^{4,5} x_2 -dependence of the solution functions for straight-crested surface waves in an unplated region and the known quantities are barred in a plated region. By taking the variations in an appropriate form^{2,3}, substituting in the variational principle and integrating with respect to depth, it has been shown^{2,3} that for large $2w$ compared to a wavelength $\lambda = 2\pi/\xi$ we obtain

$$\bar{P}(\zeta) \bar{\psi}_{,11} + \bar{Q}(\zeta) \bar{\psi}_{,1} = \bar{G}(\zeta) \bar{\psi}, \quad (2.4)$$

where

$$\hat{\psi} = \cos \bar{\alpha}_3 \bar{\psi}(x_1, t), \quad (2.5)$$

for plated region and we obtain the same form with no bars for an unplated region and \bar{P} , \bar{Q} and \bar{G} are very complicated expressions, which are too cumbersome to present here^{2,3}. Since the width ($2w$) of the array is large compared to a wavelength, we very accurately have³

$$\zeta^2 = \xi^2 + \nu^2, \quad \zeta^2 = \xi^2 + \nu^2, \quad \nu = \pi/2w, \quad \nu \ll \xi. \quad (2.6)$$

At each junction we have the continuity conditions^{2,3}

$$R(\zeta) \psi_{,1} + S(\zeta) \psi = \bar{R}(\zeta) \bar{\psi}_{,1} + \bar{S}(\zeta) \bar{\psi}, \quad \psi = \bar{\psi}, \quad (2.7)$$

where R , S , \bar{R} and \bar{S} are very complicated expressions, which are too lengthy to present here^{2,3}.

3. Transmission Equations for the Reflecting Array

In this section we briefly present the solution for a typical portion of the reflecting array containing the successive regions n , $n+1$, $n+1$, from which we obtain the desired transmission matrix. The surface wave solution functions of the approximate equations, i.e., (2.4) for the plated regions and the equivalent of (2.4) without bars for the unplated region, for the three consecutive sections may be written in the form

$$\begin{aligned}\hat{\psi}_n &= \left[C_n^R e^{i\xi(x_1 - z_n + d)} + C_n^L e^{-i\xi(x_1 - z_n + d)} \right] e^{-i\omega t}, \\ \hat{\psi}_{n+1} &= \left[\bar{C}_{n+1}^R e^{i\xi(x_1 - z_n)} + \bar{C}_{n+1}^L e^{-i\xi(x_1 - z_n)} \right] e^{-i\omega t}, \\ \hat{\psi}_{n+1} &= \left[C_{n+1}^R e^{i\xi(x_1 - z_n - \ell)} + C_{n+1}^L e^{-i\xi(x_1 - z_n - \ell)} \right] e^{-i\omega t},\end{aligned}\quad (3.1)$$

where

$$z_n = n(\ell + d), \quad (3.2)$$

and C_m^R and C_m^L denote the amplitudes of the waves traveling to the right and left, respectively, in the m th section. As in Refs 2 and 3, each solution function in (3.1) satisfies the appropriate approximate surface wave equation provided the resultant phase velocities $\bar{V} = V(\xi)$ and $V = V(\xi)$ are given by the appropriate dispersion curves in Fig. 2, where⁶

$$\omega = \xi V = \bar{\xi} \bar{V}. \quad (3.3)$$

Substituting from (3.1) into the junction conditions (2.7) at $x_1 = z_n$ and $x_1 = z_n + \ell$ and eliminating the \bar{C}_{n+1}^R and \bar{C}_{n+1}^L , we obtain the transmission equations

$$\begin{pmatrix} C_{n+1}^R \\ C_{n+1}^L \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} C_n^R \\ C_n^L \end{pmatrix}, \quad (3.4)$$

where the transmission matrix T is given by

$$\begin{aligned}T_{11} &= \{ (M^* - \bar{M}) (\bar{M}^* - M) \exp[i(\xi \ell + \xi d)] + (M^* - \bar{M}^*) \\ &\quad \times (M - \bar{M}) \exp[i(\xi d - \xi \ell)] \} [(M^* - M) (\bar{M}^* - \bar{M})]^{-1}, \\ T_{12} &= \{ (M^* - \bar{M}) (\bar{M}^* - M^*) \exp[i(\xi \ell - \xi d)] + (M^* - \bar{M}^*) \\ &\quad \times (M^* - \bar{M}) \exp[-i(\xi \ell + \xi d)] \} [(M^* - M) (\bar{M}^* - \bar{M})]^{-1}, \\ T_{21} &= \{ (M - \bar{M}) (M - \bar{M}^*) \exp[i(\xi \ell + \xi d)] + (M - \bar{M}^*) (\bar{M} - M) \\ &\quad \times \exp[i(\xi d - \xi \ell)] \} [(M^* - M) (\bar{M}^* - \bar{M})]^{-1}, \\ T_{22} &= \{ (\bar{M} - M) (\bar{M}^* - M^*) \exp[i(\xi \ell - \xi d)] + (\bar{M}^* - M) \\ &\quad \times (M^* - \bar{M}) \exp[-i(\xi \ell + \xi d)] \} [(M^* - M) (\bar{M}^* - \bar{M})]^{-1},\end{aligned}\quad (3.5)$$

where the M and \bar{M} , which are given in Refs. 2 and 3, are much too cumbersome to write here. From (3.5) it can be seen that the matrix T has the very important properties⁷

$$T_{21} = T_{12}^*, \quad T_{22} = T_{11}^*. \quad (3.6)$$

4. Continuous Representation Along the Transmission Path

The transmission equations given in (3.4) constitute a system of two linear difference equations for the

periodic array. As a solution of the linear difference equations consider

$$C_n^R = C_0^R e^{-\alpha n} e^{i n \pi}, \quad C_n^L = C_0^L e^{-\alpha n} e^{i n \pi}, \quad (4.1)$$

the substitution of which in (3.4) yields

$$\begin{aligned}C_0^R e^{-\alpha(n+1)} e^{i(n+1)\pi} &= (T_{11} C_0^R + T_{12} C_0^L) e^{-\alpha n} e^{i n \pi}, \\ C_0^L e^{-\alpha(n+1)} e^{i(n+1)\pi} &= (T_{21} C_0^R + T_{22} C_0^L) e^{-\alpha n} e^{i n \pi}.\end{aligned}\quad (4.2)$$

When the quantity $e^{-\alpha n} e^{i n \pi}$ is cancelled from both sides of (4.2) and, since $|\alpha| \ll 1$, we substitute $e^{-\alpha} = 1 - \alpha$, we obtain

$$\begin{aligned}(1 + T_{11} - \alpha) C_0^R + T_{12} C_0^L &= 0, \\ T_{21} C_0^R + (1 + T_{22} - \alpha) C_0^L &= 0.\end{aligned}\quad (4.3)$$

This system of two homogeneous linear algebraic equations in C_0^R and C_0^L yields nontrivial solutions when the determinant of the coefficients of C_0^R and C_0^L vanishes. By virtue of (3.6)₂ we may write

$$1 + T_{11} = \Delta^R + i \Delta^I, \quad 1 + T_{22} = \Delta^R - i \Delta^I, \quad (4.4)$$

which is convenient because the real parts of T_{11} and T_{22} differ slightly from -1 and enables the two roots of the quadratic determinantal equation to be written in the form

$$\alpha_{1,2} = \Delta^R \pm \sqrt{T_{12} T_{21} - (\Delta^I)^2}, \quad (4.5)$$

the substitution of each of which in (4.3), yields the respective amplitude ratios

$$C_0^{L1} = r_1 C_0^{R1}, \quad C_0^{L2} = r_2 C_0^{R2}, \quad (4.6)$$

where

$$r_{1,2} = - (i \Delta^I \mp \sqrt{T_{12} T_{21} - (\Delta^I)^2}) / T_{12}. \quad (4.7)$$

Thus, the continuous representation consists of the sum of the two eigensolutions of the linear difference equations and may be written in the form

$$\begin{aligned}C_n^R &= (C_0^{R1} e^{-\alpha_1 n} + C_0^{R2} e^{-\alpha_2 n}) e^{i n \pi}, \\ C_n^L &= (r_1 C_0^{R1} e^{-\alpha_1 n} + r_2 C_0^{R2} e^{-\alpha_2 n}) e^{i n \pi},\end{aligned}\quad (4.8)$$

where C_0^{R1} and C_0^{R2} are two arbitrary constants to be determined from suitable input conditions.

In the problem of interest there is a known input wave of amplitude \bar{C} traveling to the right incident on the left end of the array, i.e., at $n=0$, and there is no wave traveling to the left at the right end of the array, i.e., at $n=N$. Under these circumstances we have the boundary conditions

$$C_0^R = \bar{C} \quad \text{at } n=0, \quad C_N^L = 0 \quad \text{at } n=N, \quad (4.9)$$

the substitution of which in (4.8) yields

$$C_{01}^{R1} = -r_2 e^{-\alpha_2 N} \bar{C}/b, \quad C_{01}^{R2} = r_1 e^{-\alpha_1 N} \bar{C}/b, \quad (4.10)$$

where

$$b = r_1 e^{-\alpha_1 N} - r_2 e^{-\alpha_2 N}. \quad (4.11)$$

The substitution of (4.10) in (4.8) yields the solution for the problem at hand, which may be written in the form

$$C_n^R = \left[r_1 e^{-(\alpha_1 N + \alpha_2 n)} - r_2 e^{-(\alpha_2 N + \alpha_1 n)} \right] e^{in\pi} \bar{C}/b, \\ C_n^L = r_1 r_2 \left[e^{-(\alpha_1 N + \alpha_2 n)} - e^{-(\alpha_2 N + \alpha_1 n)} \right] e^{in\pi} \bar{C}/b. \quad (4.12)$$

The amplitudes of the transmitted and reflected waves may now readily be obtained from (4.12) by writing the expressions for C_N^R and C_0^L , respectively, which take the form

$$C_0^L = \frac{r_1 r_2}{b} (e^{-\alpha_1 N} - e^{-\alpha_2 N}) \bar{C}, \\ C_N^R = \frac{(r_1 - r_2)}{b} e^{-(\alpha_1 + \alpha_2)N} e^{iN\pi} \bar{C}. \quad (4.13)$$

The power reflection and power transmission coefficients R_p and T_p , respectively defined by

$$R_p = 10 \log_{10} (C_{00}^{L*} / \bar{C} \bar{C}^*), \\ T_p = 10 \log_{10} (C_N^{R*} / \bar{C} \bar{C}^*), \quad (4.14)$$

may now readily be calculated for a given configuration and a given number of strips. Such calculations have been performed for a particular configuration consisting of 200 aluminum strips 10,000 Å thick, deposited on ST-cut quartz with $d = \lambda$ and a center frequency of 74.9 MHz. The results of the calculations are plotted in Figs.3 and 4, both of which are virtually indistinguishable from the results obtained by cascading the transmission matrix^{2,3}. The amplitude and power reflection coefficients for the same geometric arrangement of strips, but for a varying number N have been calculated at the center frequency of 74.9 MHz and are plotted in Fig.5, which is virtually indistinguishable from the equivalent figure obtained by cascading the transmission matrix^{2,3}. The phase of the reflected wave for 200 strips is plotted as a function of frequency in Fig.6. The equivalent figure, i.e., Fig.13, of Ref.3 is plotted incorrectly and should have been plotted in the same way as Fig.6 of this work.

In view of the phase factors $e^{in\pi}$ in (4.8) and the forms of the wave functions in (3.1), which were used in determining the transmission matrix T , and the fact that the impedance discontinuity between the plated and unplated regions is very small in the vicinity of the fundamental resonance, the expressions for C_n^R and C_n^L without the phase factors $e^{in\pi}$ are the slowly varying (with n) amplitude coefficients of the respective continuous wave functions $e^{i\hat{\xi}x_1}$ and $e^{-i\hat{\xi}x_1}$ throughout the array, where the average wavenumber $\hat{\xi} = (\xi + \bar{\xi})/2$. Now, in order to obtain the surface wave mode shape at resonance, which is of primary interest in this work, we first write (4.12) in the form

$$C_n^R = \hat{C}_n^R e^{in\pi}, \quad C_n^L = \hat{C}_n^L e^{in\pi}, \quad (4.15)$$

then let n be a continuous variable and in place of Eqs.(4.12) with (4.15) we write

$$\hat{C}^R(x_1) = \left[r_1 e^{-\alpha_1 N} e^{-\beta_2 x_1} - r_2 e^{-\alpha_2 N} e^{-\beta_1 x_1} \right] \bar{C}/b, \\ \hat{C}^L(x_1) = r_1 r_2 \left[e^{-\alpha_1 N} e^{-\beta_2 x_1} - e^{-\alpha_2 N} e^{-\beta_1 x_1} \right] \bar{C}/b, \quad (4.16)$$

where

$$x_1 = n(\ell + d), \quad \beta_{1,2} = \alpha_{1,2}/(\ell + d) \dots \quad (4.17)$$

Clearly, Eqs.(4.16) give the aforementioned slowly varying (with x_1) amplitude coefficients of the continuous wave functions. From the foregoing statement we see that at resonance we very accurately have

$$\psi = \left[\hat{C}^R(x_1) e^{i\hat{\xi}x_1} + \hat{C}^L(x_1) e^{-i\hat{\xi}x_1} \right] e^{-i\omega t}, \quad (4.18)$$

throughout the entire array. Consequently, from (2.1), (2.5) and (4.18) the entire surface wave mode shape may be written in the form

$$u_j = \cos vx_3 \operatorname{Re} \alpha_j(x_2) \left[\hat{C}^R(x_1) e^{i\hat{\xi}x_1} + \hat{C}^L(x_1) e^{-i\hat{\xi}x_1} \right] e^{-i\omega t}, \quad (4.19)$$

the substitution of (4.16) in which yields

$$u_j = \cos vx_3 \operatorname{Re} \alpha_j(x_2) \left[r_1 e^{-\alpha_1 N} e^{-\beta_2 x_1} (e^{i\hat{\xi}x_1} + r_2 e^{-i\hat{\xi}x_1}) - r_2 e^{-\alpha_2 N} e^{-\beta_1 x_1} (e^{i\hat{\xi}x_1} + r_1 e^{-i\hat{\xi}x_1}) \right] e^{-i\omega t} \bar{C}/b, \quad (4.20)$$

which gives the surface wave mode shape at resonance throughout the entire reflecting array. In Eqs.(4.19) and (4.20) it must be remembered that while α_j goes with $\hat{\xi}$, α_j^* goes with $-\hat{\xi}$ by virtue of (2.3) and the fact that the surface wave must decay with x_2 .

Equation (4.20) may now be employed in a perturbation integral along with a biasing state to calculate a change in resonant frequency of a surface wave resonator.

For the surface wave resonator considered in Figs.3 and 4 the standing surface wave mode shapes along the transmission path with time-dependence $\cos \omega t$ and $\sin \omega t$ have been calculated at resonance and are plotted in Figs.7 and 8, respectively. It can be seen from the figures that the standing surface wave mode shape at resonance has essentially zero phase throughout the array and that the $\sin \omega t$ term can be seen at all only near the front end of the array where the mode is largest. This small $\sin \omega t$ term arises because a small amount of power is transmitted at resonance.

Acknowledgements

We wish to thank Dr. Joon Tae Song of Sung Kyun Kwan University for help with the calculations.

This work was supported in part by the Army Research Office under Contract No. DAAG 29-85-K-0075.

References

1. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., **64**, 832 (1978).
2. B.K. Sinha and H.F. Tiersten, "A Variational Analysis of the Reflection of Surface Waves by Arrays of Reflecting Strips," J. Appl. Phys., **47**, 2824 (1976).
3. B.K. Sinha and H.F. Tiersten, "An Analysis of Transverse Modes in Acoustic Surface Wave Resonators," J. Appl. Phys., **51**, 3099 (1980).
4. J.J. Campbell and W.R. Jones, "A Method for Estimating Optimal Crystal Cuts and Propagation Directions for Excitation of Piezoelectric Surface Waves," IEEE Trans. Sonics Ultrason., **SU-15**, 209 (1968).
5. B.K. Sinha and H.F. Tiersten, "Elastic and Piezoelectric Surface Waves Guided by Thin Films," J. Appl. Phys., **44**, 4831 (1973).
6. By virtue of (2.6)₄, V and \bar{V} may very accurately be taken to be the phase velocities of straight-crested waves in the plated and unplated regions, respectively.
7. References 2 and 3 incorrectly refer to the T matrix as hermitian. It is not hermitian. Instead it has the properties indicated in Eqs. (3.6).

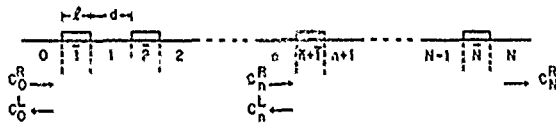


Figure 1 Schematic Diagram of a Surface-Wave Reflector

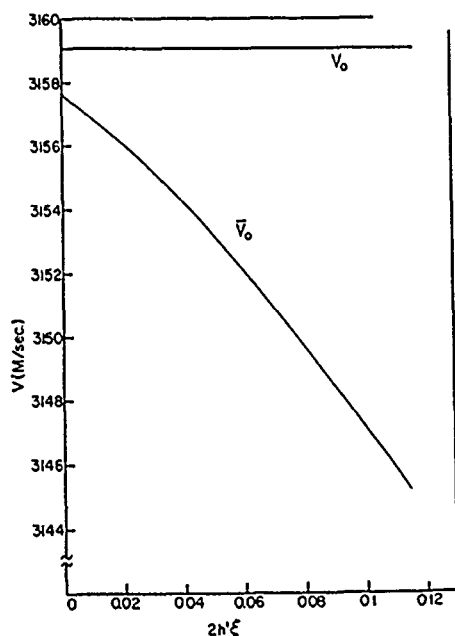


Figure 2 Lowest Straight-Crested Phase Velocity Dispersion Curve for an Aluminum Film on St-Cut Quartz

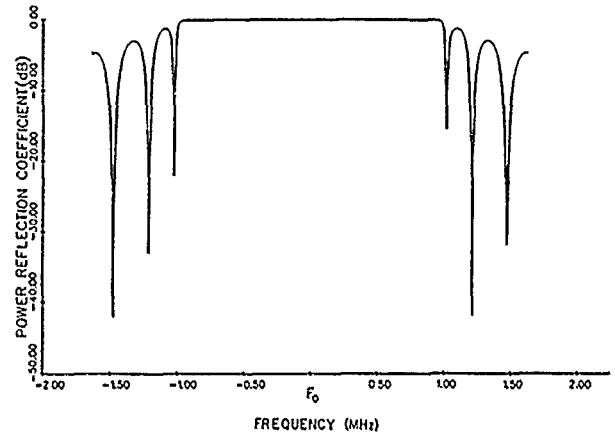


Figure 3 Power Reflection Coefficient of a Surface-Wave Reflector with 200 Aluminum Strips 10,000 Å Thick on ST-Cut Quartz at a Fundamental Frequency of 74.9 MHz. Results are for $d = l$ and $2w = 100 \lambda_R$.

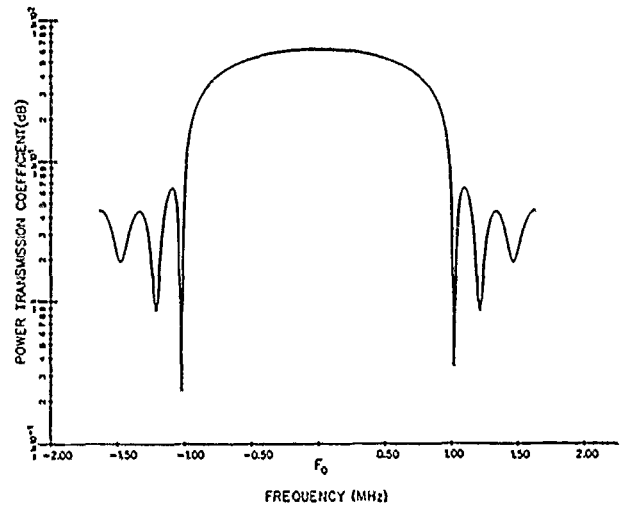


Figure 4 Power Transmission Coefficient of the Surface-Wave Reflector

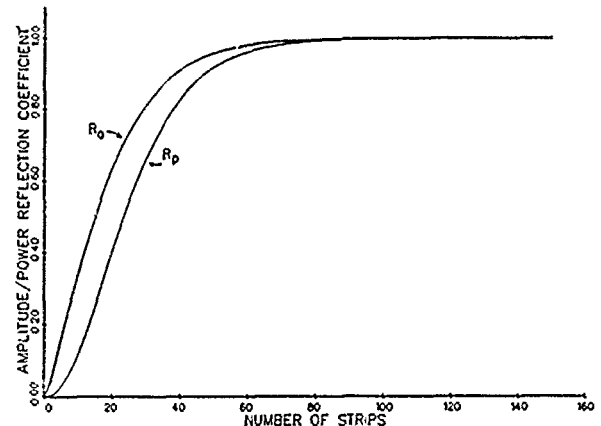


Figure 5 Amplitude (R_a) and Power (R_p) Reflection Coefficients of the Surface-Wave Reflector with 10,000 Å Thick Aluminum Strips at a Center Frequency of 74.9 MHz, as a Function of the Number of Strips

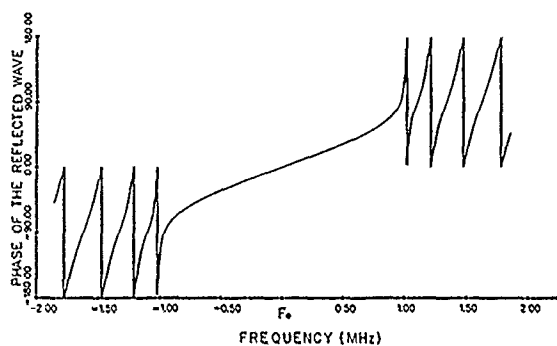


Figure 6 Phase of the Reflected Wave

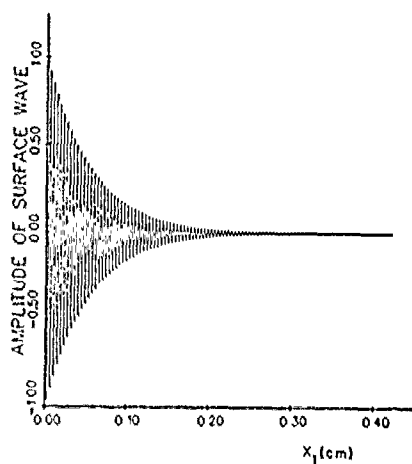


Figure 7 Standing Surface Wave Mode Shape Along Transmission Path at Resonance

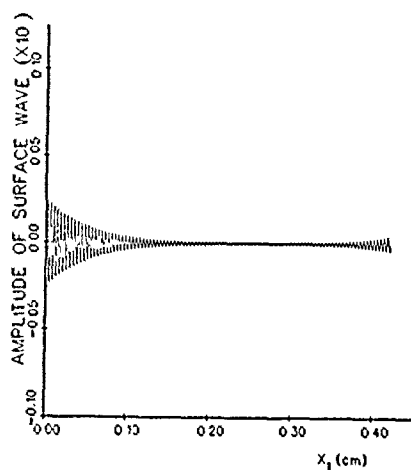


Figure 8 Out of Phase Mode Shape

AN ANALYSIS OF THE ACCELERATION SENSITIVITY OF ST-CUT QUARTZ SURFACE WAVE RESONATORS SUPPORTED ALONG THE EDGES

D.V. Shick and H.F. Tiersten
Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, New York 12180-3590

Abstract

An analysis of the acceleration sensitivity of ST-cut quartz surface wave resonators supported along rectangular edges is performed. For the case of normal acceleration the edges are assumed to be simply-supported. Although this support configuration produces flexure in the quartz substrate, which tends to produce poor acceleration sensitivity, for ST-cut quartz cylindrical flexure in the propagation direction and normal to it causes changes in frequency of opposite sign, which results in a cancellation of the changes and relatively low acceleration sensitivity. In fact by appropriate selection of the planar spanning dimensions the normal acceleration sensitivity can be made to vanish. For the case of in-plane acceleration the substrate is supported rigidly along the edges in such a way that no flexure about the thin dimension of the quartz substrate is induced. For this case the biasing deformation field is obtained using a new approximation procedure which is very accurate and convenient in perturbation integrals and for many other purposes. The calculated in-plane acceleration sensitivity turns out to be a few parts in 10^{11} per g.

1. Introduction

When a surface wave resonator is subject to cylindrical flexure the change in frequency is much greater than when it is subject to extension because flexure results in the greatest deformation near the surface of the substrate, which is exactly where the wave is located. However, for ST-cut quartz cylindrical flexure in the propagation direction and normal to it causes changes in frequency of opposite sign. Consequently, for support along all four edges of a rectangular ST-cut quartz substrate subject to normal acceleration the sensitivity can be made quite small. This has already been observed experimentally by Montress, Parker and Callera¹.

In this work an analysis of the normal acceleration sensitivity of an ST-cut quartz resonator simply supported along all four edges is performed. Simple supports are assumed in the case of flexure because this simplifies the analysis for the biasing state and we are not as interested in quantitative detail here as in a qualitative explanation of what occurs when the substrate is subject to flexure and supported along all four edges. Furthermore, it should be noted that this assumption results in an overestimate of the normal acceleration sensitivity. The biasing state is determined by solving the equations of anisotropic static flexure² which were obtained from Mindlin's plate equations³⁻⁵, and is employed in the equation for the perturbation of eigenfrequency⁶ along with an approximation to the surface wave mode shape to calculate the change in frequency. The approximation to the surface wave mode shape that is employed is now known⁷ to cause the normal acceleration sensitivity to be overestimated. Nevertheless, the resulting normal acceleration sensitivity turns out to be quite small for reasonable planar geometric ratios and in fact can be made to vanish by appropriate selection of the planar geometric ratios.

In addition, an analysis of the in-plane acceleration sensitivity of an ST-cut quartz resonator supported rigidly along all four edges is performed. Since for this support configuration no flexure about the thin dimension of the quartz substrate is induced by in-plane acceleration, this should be an excellent configuration for low in-plane acceleration sensitivity. The extensional biasing state is determined by means of a new variational approximation procedure, which is very accurate and results in convenient forms for use in the perturbation integrals. The extensional biasing state is employed in the perturbation equation along with the aforementioned approximation to the mode shape to calculate the change in frequency. The resulting in-plane acceleration sensitivity is a few parts in 10^{11} per g.

2. Perturbation Equations

For purely elastic nonlinearities the equation for the first perturbation of the eigenvalue obtained from the perturbation analysis⁶ mentioned in the Introduction may be written in the form

$$\Delta_{\mu} = H_{\mu} / 2\omega_{\mu}, \quad \omega = \omega - \Delta_{\mu}, \quad (2.1)$$

where ω_{μ} and ω are the unperturbed and perturbed eigenfrequencies, respectively, and

$$H_{\mu} = - \int_V \tilde{K}_{LY}^n g_{Y,L}^{\mu} dV, \quad (2.2)$$

where V is the undeformed volume of the piezoelectric plate. In (2.2) g_Y^{μ} denotes the normalized mechanical displacement vector, and \tilde{K}_{LY}^n denotes the portion of the Piola-Kirchhoff stress tensor resulting from the biasing state in the presence of the g_Y^{μ} , and is given by

$$\tilde{K}_{LY}^n = \hat{c}_{LYM\alpha} g_{M,\alpha}^{\mu} \quad (2.3)$$

where

$$\hat{c}_{LYM\alpha} = T_{LM}^1 \delta_{Y\alpha} + c_{3LYM\alpha KN} E_{KN}^1 + c_{2LYKM} w_{\alpha,K} + c_{2LKM\alpha} w_{Y,K}, \quad (2.4)$$

and

$$T_{LM}^1 = c_{2LMKN} E_{KN}^1, \quad E_{KN}^1 = \frac{1}{2} (w_{K,N} + w_{N,K}). \quad (2.5)$$

The quantities T_{LM}^1 , E_{KN}^1 and w_K denote the static biasing stress, strain and displacement field, respectively. Thus, in this description the present position \underline{y} is related to the reference position \underline{x} by

$$\underline{y}(\underline{x}_L, t) = \underline{x} + \underline{w}(\underline{x}_L) + \underline{u}(\underline{x}_L, t). \quad (2.6)$$

The coefficients c_{2LMKN} and $c_{3LYM\alpha KN}$ denote the second and third order elastic constants, respectively.

The normalized eigensolution g_Y^μ and \hat{F}^μ is defined by

$$g_Y^\mu = \frac{u_Y^\mu}{N_\mu}, \quad \hat{F}^\mu = \frac{\tilde{\phi}^\mu}{N_\mu}, \quad N_\mu^2 = \int_V \rho u_Y^\mu u_Y^\mu dV, \quad (2.7)$$

where u_Y^μ and $\tilde{\phi}^\mu$ are the mechanical displacement and electric potential, respectively, which satisfy the equations of linear piezoelectricity

$$\tilde{K}_{LY}^\ell = c_{LYM\alpha} u_{\alpha,M} + e_{MLY} \tilde{\phi}_{,M}, \quad (2.8)$$

$$\tilde{D}_L^\ell = e_{LMY} u_{Y,M} - e_{LMY} \tilde{\phi}_{,M}, \quad (2.9)$$

subject to the appropriate boundary conditions, and ρ is the mass density. Equations (2.8) are the linear piezoelectric constitutive relations and (2.9) are the stress equations of motion and charge equation of electrostatics, respectively. The upper cycle notation for many dynamic variables and the capital Latin and lower case Greek index notation is being employed for consistency with Ref. 6, as is the remainder of the notation in this section.

The substitution of (2.3) in (2.2) yields

$$H_\mu = - \int_V \hat{c}_{LYM\alpha} g_Y^\mu g_{Y,L}^\mu dV. \quad (2.10)$$

Since g_Y^μ denotes the normalized surface wave mode shape and from (2.4) $\hat{c}_{LYM\alpha}$ depends on the biasing state, H_μ can be evaluated when the surface wave mode shape and biasing state are known.

3. Flexure of Simply-Supported Rectangular ST-Cut Quartz Plate

It has been shown that the equation of anisotropic static flexure with X_2 normal to the major surfaces can be written in the form

$$K_{AB,AB}^{(1)} + F_{B,B}^{(1)} + F_2^{(0)} = 0, \quad (3.1)$$

where we have introduced the convention that A, B, C, D take the values 1 and 3 and skip 2 and

$$K_{AB}^{(1)} = \int_{-h}^h X_2 K_{AB} dX_2, \quad F_B^{(1)} = h[K_{2B}(h) + K_{2B}(-h)], \quad (3.2)$$

and in this instance

$$F_B^{(1)} = 0, \quad F_2^{(0)} = K_{22}(h) - K_{22}(-h) - 2hpa_2 = -2hpa_2, \quad (3.3)$$

since K_{2K} vanishes on the major surfaces and where a_2 is the acceleration in the X_2 -direction. From Eqs. (4.25) of Ref. 2 the constitutive equations for the stress-resultants take the form

$$K_{AB}^{(1)} = \frac{2}{3} h^3 \gamma_{ABCD} E_{CD}^{(1)}, \quad (3.4)$$

where Voigt's anisotropic plate elastic constants are given by

$$\gamma_{RS} = c_{RS} - c_{RW} c_{WV} c_{VS}^{-1}, \quad R, S = 1, 3, 5; \quad W, V = 2, 4, 6, \quad (3.5)$$

in the compressed notation, and where we have introduced the scheme shown. The plate strains $E_{\alpha\beta}^{(1)}$ in (3.4) are given by

$$E_{CD}^{(1)} = \frac{1}{2} (w_{C,D}^{(1)} + w_{D,C}^{(1)}), \quad (3.6)$$

and from the relaxation of the stress resultants $K_{2L}^{(1)}$, the vanishing of the plate shear strains $E_{2A}^{(0)}$ and (3.6) we have the respective relations

$$E_W^{(1)} = -\frac{1}{c_{WV} c_{VS}} E_S^{(1)}, \quad w_A^{(1)} = -w_{2,A}^{(0)}, \quad E_{CD}^{(1)} = -w_{2,AB}^{(0)}, \quad (3.7)$$

which, respectively, are given in Eqs. (4.24), (3.34) and (3.37) of Ref. 2. Since from Eqs. (3.31) of Ref. 9 we have $E_{KL}^{(0)} = 0$, when the plate deflection $w_2^{(0)}$ has been found, we know the three-dimensional strain field from the relation

$$E_{KL} = \frac{1}{2} (w_{L,K} + w_{K,L}) = X_2 E_{KL}^{(1)}. \quad (3.8)$$

as in Eqs. (3.32) of Ref. 9.

Although we now have the plate strains from (3.8), we cannot yet determine the $\hat{c}_{LYM\alpha}$ from (2.4) because, as noted in Ref. 9, we need the displacement gradients $w_{L,K}$ or equivalently the rotations Ω_{KL} , which are defined by

$$\Omega_{KL} = \frac{1}{2} (w_{L,K} - w_{K,L}). \quad (3.9)$$

As in Ref. 9, we require that the associated plate rotations that accompany the plate strains that arise from the relaxation of the plate stress resultants $K_{2L}^{(1)}$ satisfy the appropriate three-dimensional rotation gradient-strain gradient relations, i.e.,

$$\Omega_{KL,M} = E_{ML,K} - E_{MK,L}. \quad (3.10)$$

From (3.22) and (3.23) of Ref. 9 we have the expressions for the plate rotations

$$\Omega_{AB}^{(n)} = \frac{1}{2} (w_{B,A}^{(n)} - w_{A,B}^{(n)}),$$

$$\Omega_{2A}^{(n)} = \frac{1}{2} [(n+1)w_A^{(n+1)} - w_{2,A}^{(n)}]. \quad (3.11)$$

From (3.33) of Ref. 9, we have

$$\Omega_{13}^{(0)} = \text{arbitrary constant} = 0, \quad \Omega_{13}^{(1)} = 0, \quad (3.12)$$

and from (3.35) and (3.34), respectively, of Ref. 9, we have

$$\Omega_{2A}^{(0)} = -w_{2,A}^{(0)}, \quad \Omega_{2A}^{(1)} = E_{2A}^{(1)}, \quad (3.13)$$

which with (3.24) of Ref. 9 enables us to write

$$\Omega_{2A} = -w_{2,A}^{(0)} + X_2 E_{2A}^{(1)}, \quad \Omega_{13} = 0, \quad (3.14)$$

which is the same as (3.36) of Ref. 9. Now, from (3.8), (3.14) and

$$w_{K,L} = E_{KL} + \Omega_{LK}, \quad (3.15)$$

we have the desired three-dimensional displacement gradients when the plate deflection $w_2^{(0)}$ has been found.

A plan view and cross-section of the simply-supported rectangular plate is shown in Fig.1 along with the coordinate system. The substitution of (3.3), (3.4) and (3.7)₃ in (3.1) yields the equilibrium equation for flexure of the thin plate in the form

$$\frac{2}{3} h^3 \gamma_{ABCD} w_{2,CDAB}^{(0)} + 2h\rho a_2 = 0. \quad (3.16)$$

Since the plate is simply-supported along the edges, the boundary conditions take the form

$$\begin{aligned} w_2^{(0)} = 0, \quad K_{11}^{(1)} = 0 \quad \text{at } X_1 = \pm a, \\ w_2^{(0)} = 0, \quad K_{33}^{(1)} = 0 \quad \text{at } X_3 = \pm b. \end{aligned} \quad (3.17)$$

Since for singly-rotated Y-cut quartz plates (such as the ST-cut) $\gamma_{15} = \gamma_{35} = 0$, we can obtain a series solution for the simply-supported plate in the form

$$w_2^{(0)} = \sum_m \sum_n A_{mn} \cos \alpha_m X_1 \cos \kappa_n X_3, \quad (3.18)$$

which satisfies (3.17) provided

$$\alpha_m = m\pi/2a, \quad \kappa_n = n\pi/2b, \quad m \text{ and } n \text{ odd}. \quad (3.19)$$

Substituting from (3.18) into (3.16) and employing the orthogonality of the trigonometric functions, we obtain

$$A_{mn} = \frac{48 \rho a_2}{h^2 m n \pi^2 R_{mn}}, \quad (3.20)$$

where

$$R_{mn} = -[\gamma_{11} \alpha_m^4 + \gamma_{33} \kappa_n^4 + (2\gamma_{13} + 4\gamma_{55}) \alpha_m^2 \kappa_n^2]. \quad (3.21)$$

This yields $w_2^{(0)}$, from which we can obtain the biasing $w_{K,L}$ from Eqs. (3.7)₃, (3.7)₁, (3.14) and (3.15).

4. Extension of Rectangular ST-Cut Quartz Plate Rigidly Supported Along the Edges

Since the plate is thin, the equations of anisotropic plane stress apply, which may be written in the form

$$K_{AB,A} - \tau_{a_B} = 0, \quad (4.1)$$

where a_B denotes the acceleration in the plane of the plate and the anisotropic constitutive equations for plane stress may be written in the form

$$K_{AB} = \gamma_{ABCD} E_{CD}, \quad E_{CD} = \frac{1}{2} (w_{C,D} + w_{D,C}), \quad (4.2)$$

where γ_{ABCD} are Voigt's anisotropic plate elastic constants given in (3.5)₁. From the relaxation of the stresses K_{2L} , we have

$$E_W = -c_{WV}^{-1} c_{VS} E_S, \quad (4.3)$$

which enables us to obtain all the strains E_{KL} when a solution is known as well as the rotation Ω_{13} from the relation

$$\Omega_{13} = \frac{1}{2} (w_{3,1} - w_{1,3}). \quad (4.4)$$

However, the rotations Ω_{2A} cannot be obtained from a solution because of the relaxation of the K_{2L} .

Nevertheless, it has been shown in Eqs. (3.28) of Ref. 9 that the rotations Ω_{2A} may be obtained a posteriori by solving the equations

$$\Omega_{2A,B} = -E_{B2,A}, \quad (4.5)$$

which are a consequence of (3.10), and now after solving (4.5) and employing

$$w_{K,L} = E_{KL} + \Omega_{LK}, \quad (4.6)$$

we have the required three-dimensional displacement gradients for use in (2.4).

A plan view and cross-section of the plate is shown in Fig.2 along with the coordinate system. Since the plate has all four edges fixed, we have the boundary conditions

$$\begin{aligned} w_1 = w_3 = 0 \quad \text{at } X_1 = \pm a, \quad |X_3| \leq b, \\ w_1 = w_3 = 0 \quad \text{at } X_3 = \pm b, \quad |X_1| \leq a. \end{aligned} \quad (4.7)$$

In obtaining the solution to this problem we first transform the inhomogeneities from the differential equations, which consist of (4.1) with (4.2), into the boundary conditions (4.7) by writing

$$w_1 = A_1 (X_1^2 - a^2) + \hat{w}_1, \quad w_3 = A_3 (X_3^2 - b^2) + \hat{w}_3, \quad (4.8)$$

which when substituted into (4.1) with (4.2) yields

$$\begin{aligned} \gamma_{11} \hat{w}_{1,11} + (\gamma_{13} + \gamma_{55}) \hat{w}_{3,13} + \gamma_{55} \hat{w}_{1,33} = 0, \\ \gamma_{55} \hat{w}_{3,11} + (\gamma_{13} + \gamma_{55}) \hat{w}_{1,13} + \gamma_{33} \hat{w}_{3,33} = 0, \end{aligned} \quad (4.9)$$

since for ST-cut quartz $\gamma_{15} = \gamma_{35} = 0$ and A_1 and A_3 have been selected as

$$A_1 = \rho a_1 / 2\gamma_{11}, \quad A_3 = \rho a_3 / 2\gamma_{33}. \quad (4.10)$$

The further substitution of (4.8) into (4.7) yields the edge conditions

$$\hat{w}_1 = 0, \quad \hat{w}_3 = -A_3 (X_3^2 - b^2) \quad \text{at } X_1 = \pm a, \quad |X_3| \leq b, \quad (4.11)$$

$$\hat{w}_1 = -A_1 (X_1^2 - a^2), \quad \hat{w}_3 = 0 \quad \text{at } X_3 = \pm b, \quad |X_1| \leq a. \quad (4.12)$$

Since we now consider acceleration in the X_1 -direction only, we have $a_3 = 0$, which with (4.10)₂ yields $A_3 = 0$ and in place of (4.11) we have

$$\hat{w}_1 = \hat{w}_3 = 0 \quad \text{at } X_1 = \pm a, \quad |X_3| \leq b, \quad (4.13)$$

while (4.12) still holds. As a solution of (4.9) consider the finite sum

$$\hat{w}_1 = \sum_m \beta_{1m} e^{i\eta_m X_3} e^{i\nu_m X_1}, \quad \hat{w}_3 = \sum_m \beta_{3m} e^{i\eta_m X_3} e^{i\nu_m X_1}, \quad (4.14)$$

where $\nu_m = m\pi/2a$ and m takes all + and - integers considered in the sum. The substitution of (4.14) into (4.9) yields

$$\begin{aligned} (\nu_m^2 \gamma_{11} + \eta_m^2 \gamma_{55}) \beta_{1m} + (\gamma_{13} + \gamma_{55}) \nu_m \eta_m \beta_{3m} = 0, \\ (\gamma_{13} + \gamma_{55}) \nu_m \eta_m \beta_{1m} + (\gamma_{55} \nu_m^2 + \gamma_{33} \eta_m^2) \beta_{3m} = 0, \end{aligned} \quad (4.15)$$

for each m . Equations (4.15) constitute a system of two linear homogeneous algebraic equations in β_{1m} and

β_{3m} , which yields nontrivial solutions when the determinant of the coefficients vanishes. If we define $\eta_m = h\nu_m$, then the aforementioned vanishing determinant yields

$$\gamma_{55}\gamma_{33}h^4 + (\gamma_{11}\gamma_{33} - \gamma_{13}^2 - 2\gamma_{13}\gamma_{55})h^2 + \gamma_{11}\gamma_{55} = 0, \quad (4.16)$$

which has four roots $h^{(p)}$ ($p=1,2,3,4$), each of which yields a set of amplitude ratios $\beta_1^{(p)}:\beta_3^{(p)}$, which are independent of m . Hence, as a solution of the boundary value problem we take

$$\begin{aligned} \hat{w}_1 &= \sum_m \sum_{p=1}^4 D_p^m \beta_1^{(p)} e^{i(\eta_m^p X_3 + \nu_m X_1)}, \\ \hat{w}_3 &= \sum_m \sum_{p=1}^4 D_p^m \beta_3^{(p)} e^{i(\eta_m^p X_3 + \nu_m X_1)}, \end{aligned} \quad (4.17)$$

where the D_p^m are amplitude coefficients still to be determined and $\eta_m^p = h^{(p)}\nu_m$. By requiring (4.12)₂ to be satisfied for each m , we can eliminate two of the D_p^m for each m and thereby reduce the number of unknown coefficients by a factor of two. In this way we obtain

$$D_1^m = g_{13}^{(m)} D_3^m + g_{14}^{(m)} D_4^m, \quad D_2^m = g_{23}^{(m)} D_3^m + g_{24}^{(m)} D_4^m, \quad (4.18)$$

and for brevity we do not bother to write the expressions for the $g_{rs}^{(m)}$ here¹⁰. Substituting from (4.18) into (4.17), we obtain the reduced form

$$\begin{aligned} \hat{w}_A &= \sum_m \left[\left(g_{13}^{(m)} \beta_A^{(1)} e^{i\eta_m^{(1)} X_3} + g_{23}^{(m)} \beta_A^{(2)} e^{i\eta_m^{(2)} X_3} \right. \right. \\ &\quad \left. \left. + \beta_A^{(3)} e^{i\eta_m^{(3)} X_3} \right) D_3^m + \left(g_{14}^{(m)} \beta_A^{(1)} e^{i\eta_m^{(1)} X_3} \right. \right. \\ &\quad \left. \left. + g_{24}^{(m)} \beta_A^{(2)} e^{i\eta_m^{(2)} X_3} + \beta_A^{(4)} e^{i\eta_m^{(4)} X_3} \right) D_4^m \right] e^{i\nu_m X_1}, \end{aligned} \quad (4.19)$$

for the solution function.

The solution functions in (4.19) satisfy the differential equation and boundary conditions (4.12)₂, but they do not satisfy the boundary conditions (4.12)₁ and (4.13). We satisfy these latter conditions approximately by means of a variational approximation procedure. Since the solution functions satisfy the differential equations exactly¹¹, all that remains of the appropriate variational principle in which all conditions are unconstrained¹² is

$$\int_S N_A (w_B - \bar{w}_B) \delta K_{AB} dS = 0, \quad (4.20)$$

where N_A is the outwardly directed unit normal to the edges and \bar{w}_B is the prescribed edge displacement (here $\bar{w}_B = 0$). Substituting from (4.8) with $a_3 = 0$ into (4.20), we obtain

$$\begin{aligned} & - \int_{-a}^a \left[(A_1(X_1^2 - a^2) + \hat{w}_1) \delta \hat{K}_{13} \right]_{X_3=-b} dX_1 + \int_{-b}^b [\hat{w}_1 \delta \hat{K}_{11} \\ & + \hat{w}_3 \delta \hat{K}_{13}]_{X_1=a} dX_3 + \int_{-a}^a \left[(A_1(X_1^2 - a^2) + \hat{w}_1) \delta \hat{K}_{13} \right]_{X_3=b} dX_1 \\ & - \int_{-b}^b [\hat{w}_1 \delta \hat{K}_{11} + \hat{w}_3 \delta \hat{K}_{13}]_{X_1=-a} dX_1 = 0, \end{aligned} \quad (4.21)$$

since the variation of prescribed quantities vanishes. Substituting from (4.19) into (4.21) and performing the integrations, we obtain

$$\begin{aligned} & \sum_m \sum_n (a_{33}^{(mn)} D_3^m \delta D_3^n + a_{34}^{(mn)} D_3^m \delta D_4^n + a_{43}^{(mn)} D_4^m \delta D_3^n \\ & + a_{44}^{(mn)} D_4^m \delta D_4^n) + \sum_n (b_3^{(n)} \delta D_3^n + b_4^{(n)} \delta D_4^n) = 0, \end{aligned} \quad (4.22)$$

where the expressions for the $a_{rs}^{(mn)}$ and $b_s^{(n)}$ are too lengthy to present here¹⁰. Since the variations δD_3^n and δD_4^n are arbitrary, we have

$$\begin{aligned} & \sum_m (a_{33}^{(mn)} D_3^m + a_{43}^{(mn)} D_4^m) = -b_3^n, \\ & \sum_m (a_{34}^{(mn)} D_3^m + a_{44}^{(mn)} D_4^m) = -b_4^n, \end{aligned} \quad (4.23)$$

which constitute 2N inhomogeneous linear algebraic equations in the 2N unknowns D_3^m and D_4^m . Convergence is determined by increasing the number N and comparing the solution for N with that for N-1.

For $a_3 \neq 0$, set $a_1 = 0$ and repeat procedure interchanging 3 and 1.

5. Resonant Surface Wave Mode Shape

In this section we present an approximate continuous representation of the surface wave mode shape in resonators with grooved reflectors. Although the depth dependence and transverse mode shape are essentially exact, the mode shape along the transmission path is just an approximate guess and does not use the analysis presented in the previous paper⁷. The reason for this is that we did not want this work to wait for that work⁷ to be complete and, more importantly, we presently have not completed the transmission analysis for the grooved array and did not want this work to wait for that, which will take some time. Furthermore, simplifying assumptions made in Sec. 3 tend to overestimate the normal acceleration sensitivity anyway and, as already noted, this work is not intended to be that accurate but only to give an understanding of what is going on as early as possible.

The straight-crested surface wave displacement field may be written in the known form^{13,14}

$$\hat{u}_j = \sum_{m=1}^4 c^{(m)} A_j^{(m)} e^{i\beta_m X_2} e^{i\xi(X_1 - Vt)}. \quad (5.1)$$

A plan view of the resonator showing the reflecting arrays of grooves, the coordinate system and the associated planar geometry is given in Fig. 3. The

variable-crested surface wave with variable amplitude along the transmission path is approximated by

$$u_j = \cos \frac{\pi X_3}{2w} f(X_1) \hat{u}_j, \quad (5.2)$$

where we assume that

$$f(X_1) = 1, \quad -s < X_1 < s, \\ f(X_1) = e^{-A(X_1-s)^2}, \quad s < |X_1| < \ell, \quad (5.3)$$

which means that we fit the mode shape along the transmission path in the reflecting array to a Gaussian. We now choose A so that $f(\ell) = \delta$ (small quantity assumed known). In order to simplify the perturbation integrals we fit the Gaussian to a cosine function and decaying exponential matched at the inflection point \bar{X}_1 of the Gaussian. To this end we write

$$f(X_1) \approx \cos B(X_1 - s), \quad s < X_1 < \bar{X}_1, \\ \approx C e^{-D(X_1-s)}, \quad \bar{X}_1 < X_1 < \ell. \quad (5.4)$$

Now, carrying out the aforementioned fitting, we obtain

$$A = -\frac{\log \delta}{(\ell - s)^2}, \quad B = A^{\frac{1}{2}} \cos^{-1} e^{-\frac{1}{2}}, \quad C = e^{-D(\ell-s)}, \\ D = \frac{\frac{1}{2} + \log \delta}{\bar{X}_1 - \ell}, \quad (5.5)$$

which gives us the surface wave mode shape at resonance that we can use in the perturbation integral along with the results of either Sec.3 or 4. A comparison of the assumed mode shape along the transmission path in the reflecting array and the actual calculated mode shape from Ref. 7 is given in Fig.4. It can be seen from the figure that the assumed mode shape overestimates the effect of the mode throughout the reflecting array by quite a bit. Hence, the calculated acceleration sensitivity will be an overestimate, but not by as much as indicated in Fig.4 because of the importance of the constant region $|X_1| < s$, in which the assumed mode shape is correct.

6. Acceleration Sensitivity

From Sec.5 we now know g_y^u and from Secs.3 and 4, respectively, we know $\hat{c}_{LYM\alpha}$ for normal acceleration, i.e., for flexure with simply-supported edges, and in-plane acceleration, i.e., extension with fixed edges. Hence, we can now evaluate H_μ in Eq. (2.10) for both cases. Such calculations have been performed using the known values of the second order¹⁵ and third order¹⁶ elastic constants of quartz for the case of normal acceleration, i.e., flexure with simple supports, and the acceleration sensitivity is plotted in Fig.5 as a function of the planar aspect ratio a/b . It can be seen from the figure that the acceleration sensitivity goes through zero for a value of a/b of about 4.75. Since because of the simplifying assumption of simple supports and the overevaluation of the surface wave mode shape, whenever the acceleration sensitivity is non-zero, it is rather appreciably overestimated. Consequently, the actual curve (for non-simple supports) will be significantly shallower than the one shown in Fig.5.

Calculations have been performed for the case of in-plane acceleration also, i.e., extension with fixed edges, and the acceleration sensitivity is given in Table I. For the case $a=b=5$ mm the calculated sensitivities are given as a function of N, where $2N$ is the number of terms in the series, to indicate the convergence of the new approximation technique. It can be seen from the table that the acceleration sensitivity converges very rapidly. Converged results are presented for a few other geometries also. It is clear from the table that the calculated in-plane acceleration sensitivity for fixed edges is a few parts in 10^{11} per g for rigidly held edges. This acceleration sensitivity is somewhat overestimated also because of the overestimate of the surface wave mode shape along the transmission path shown in Fig.4.

Acknowledgements

This work was supported in part by the Army Research Office under Contract No. DAAG-29-85-K-0075.

References

1. G.K. Montress, T.E. Parker and J. Callera, "A Miniature Hybrid Circuit SAW Oscillator Using an All Quartz Packaged Resonator," 1985 Ultrasonics Symposium Proceedings, IEEE Cat. No. 85 CH2209-5, Institute of Electrical and Electronics Engineers, New York, 277 (1985).
2. H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Doubly-Rotated Quartz Thickness-Mode Resonators," J. Appl. Phys., 50, 8038 (1979).
3. R.D. Mindlin, "An Introduction to the Mathematical Theory of the Vibration of Elastic Plates," U.S. Army Signal Corps Eng. Lab., Fort Monmouth, New Jersey (1955). Signal Corps Contract DA-36-03956-56772.
4. R.D. Mindlin, "High Frequency Vibrations of Crystal Plates," Quart. Appl. Math., 19, 51 (1961).
5. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.13.
6. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 832 (1978).
7. H.F. Tiersten and D.V. Shick, "On a Continuous Representation of the Mode Shape in Acoustic Surface Wave Resonators," Proceedings of the 40th Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey and Institute of Electrical and Electronics Engineers, New York, these proceedings (1986).
8. Ref.2, Eq.(3.39).
9. D.S. Stevens, H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Contoured AT-Cut Quartz Crystal Resonators," J. Appl. Phys., 54, 1709 (1983).
10. For more detail see D.V. Shick and H.F. Tiersten, "On the Acceleration Sensitivity of Rectangular ST-Cut Quartz Surface Wave Resonators Supported Along All Four Edges," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.
11. In this particular instance not only are the differential equations satisfied exactly, but certain of the boundary conditions, namely $(4.12)_2$, are satisfied exactly also. It should be noted that the approximation procedure being applied to this

relatively simple anisotropic plane stress problem treated here has far wider applications than indicated in this work.

12. Ref.5, Eq.(6.44).
13. J.J. Campbell and W.R. Jones, "A Method for Estimating Optimal Crystal Cuts and Propagation Directions for Excitation of Piezoelectric Surface Waves," IEEE Trans. Sonics Ultrason., SU-15, 209 (1968).
14. B.K. Sinha and H.F. Tiersten, "Elastic and Piezoelectric Surface Waves Guided by Thin Films," J. Appl. Phys., 44, 4830 (1973). Sec.II.
15. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 110, 1060 (1958).
16. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

TABLE I

IN-PLANE ACCELERATION SENSITIVITY

	$a_1 \neq 0, a_3 = 0$	$a_1 = 0, a_3 \neq 0$	
N	$\Delta f/f$ per g	$\Delta f/f$ per g	
2	0.572×10^{-11}	0.619×10^{-11}	$a = b = 5 \text{ mm}$
4	0.909	0.833	
6	4.402	2.351	
8	1.918	5.395	
10	1.337	2.608	
12	1.295	2.392	
14	1.294	2.399	
16	1.2938	2.416	
18	1.2938	2.4157	
20	1.2938	2.4157	
20	2.609	4.605	$a = 5 \text{ mm},$ $b = 10 \text{ mm}$
20	3.691	3.425	$a = 10 \text{ mm},$ $b = 5 \text{ mm}$

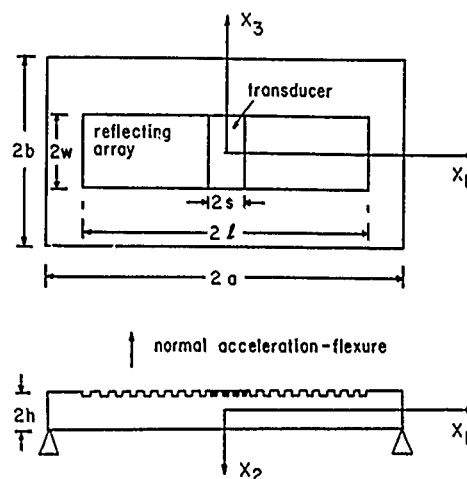


Figure 1 Plan View and Cross-Section of Simply-Supported Rectangular ST-Cut Quartz Plate Subject to Flexure

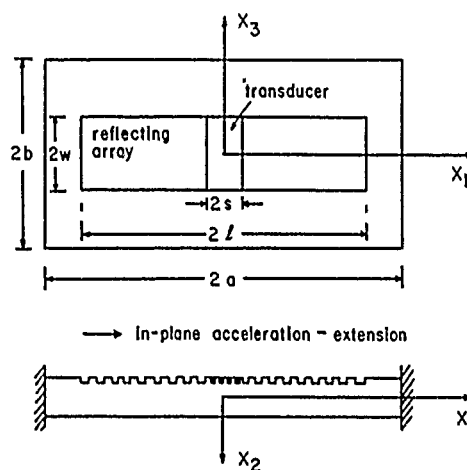


Figure 2 Plan View and Cross-Section of Rectangular ST-Cut Quartz Plate Rigidly Supported at the Edges and Subject to Extension

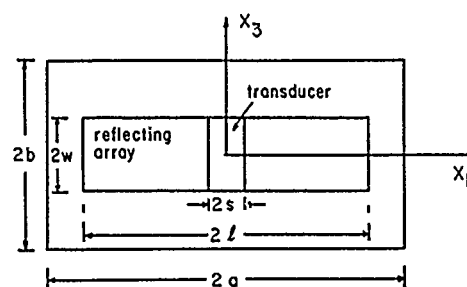


Figure 3 Plan View of Surface Wave Resonator

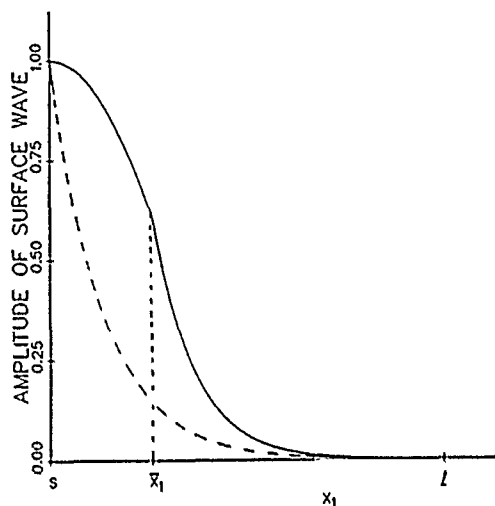


Figure 4 Comparison of Assumed and Calculated Surface Wave Mode Shape Along the Transmission Path. The calculated curve is from Ref.7.

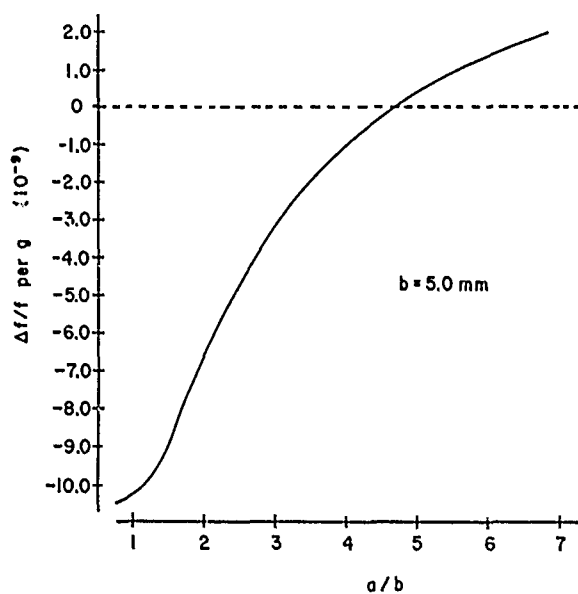


Figure 5 Calculated Normal Acceleration Sensitivity Versus Planar Aspect Ratio a/b

LOW LOSS, HIGHLY STABLE SAW DEVICES ON QUARTZ

T N Oliver, D E Bower and J Dowsett
STC Components Limited
Quartz Crystal Unit
Edinburgh Way
Harlow
Essex
United Kingdom

ABSTRACT

As the application of surface acoustic wave (SAW) devices becomes more widespread the need has arisen for improvements in insertion loss and temperature stability at higher operating frequencies. This paper reviews these areas of SAW technology and presents results on a low insertion loss structure based upon the Single Phase Unidirectional Transducer (SPUDT). This approach is shown to give devices with excellent performance whilst design, fabrication and application remain relatively straight forward.

ST-X quartz and similar cuts are selected primarily for high temperature stability requirements. Results of work on LST quartz are presented and show good resonator and filter performance indicating the possibility of application to low loss, multipole resonator, narrow band filtering applications.

BACKGROUND

Conventional, bi-directional SAW transversal filters are generally operated in a deliberately mis-matched configuration which maintains spurious internal reflections at a low level. Typically such devices will have an insertion loss in the range 12-30 dB (Fig. 1). Lower insertion losses are required for front-end filtering, where high losses degrade the system noise figure, and for low power consumption where the power requirements of gain stages need to be kept to a minimum. Typical low power applications are mobile radio and submerged repeaters for digital data transmission systems.

For very narrow band filtering SAW resonator filters with fractional bandwidths less than 0.05% have the advantage of low insertion loss when matched. Generally wider bandwidths are needed and various techniques have been proposed for achieving unidirectional transducer operation and hence low insertion loss. These include:

- Multi-strip coupler configurations [1]
- Three phase unidirectional transducers [2]
- Group-type unidirectional transducers [3]
- Single phase unidirectional transducers [4]
- Interdigitated, interdigital transducers [5]

In selecting the most suitable approach for this work the following design considerations were made:

- Optimum insertion loss
- Suitability for quartz
- Ease of design
- Ease of fabrication
- Ease of application

As a result, two different approaches which are essentially bandwidth dependent have been investigated.

Coupled resonator filters for fractional bandwidths less than 0.1%.

Group-type single phase unidirectional transducers as proposed by Lewis [6] for fractional bandwidths upto 1%.

These designs are fabricated on YX 39° quartz where the temperature performance is similar to ST-X quartz, the 39° cut being chosen simply to relocate the turnover temperature. Several other quartz cuts have been proposed which give improved temperature stability (measured as the temperature coefficient of delay or TCD) although other factors such as acoustic wave velocity, beam steering, spurious modes and diffraction effects need to be considered. Improved stability of uncompensated oscillators and narrow band filters leads to a reduction in system complexity as well as better technical performance.

Recently, Shimizu [7] has proposed a cut of quartz which utilises a leaky surface wave at an orientation where the true Rayleigh wave is well suppressed. This cut, named LST, has the advantages of small temperature coefficient and high acoustic wave velocity. In this report both resonators and filters are demonstrated on this cut.

INSERTION LOSSCoupled Resonator Filters

SAW resonators with unmatched losses in the range 7 to 15 dB when matched have resultant losses as low as 1 to 2 dB. However, the very narrow bandwidths are of little use in most filter applications. Much work has been reported (e.g. [8]) on coupling resonators

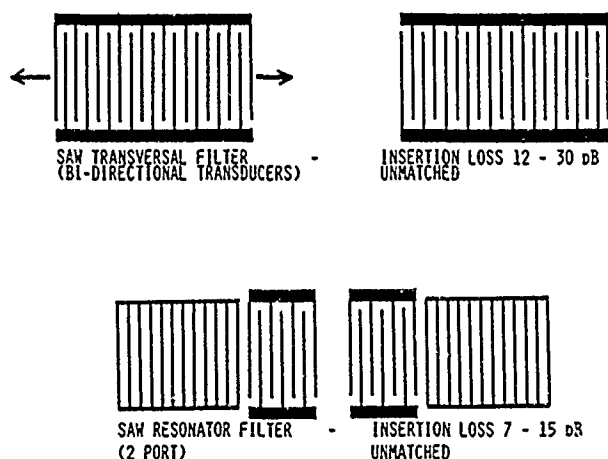


Figure 1 - Conventional SAW Structures

of similar but offset centre frequencies to achieve wider bandwidths whilst retaining low insertion loss. Such an arrangement is illustrated (Fig. 2) and shows a pair of 2-port resonators which can be coupled on the same chip or separately in individual packages.

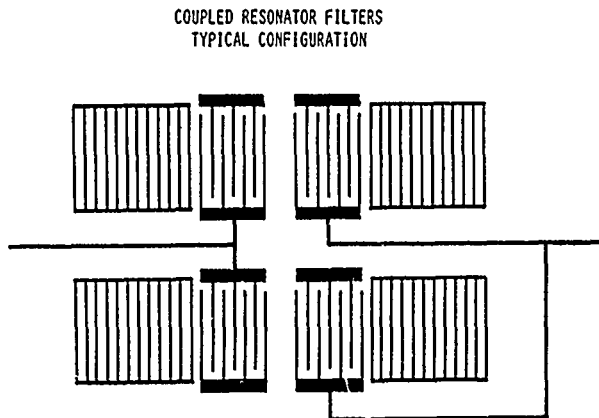


Figure 2 - Coupled Resonators - Layout

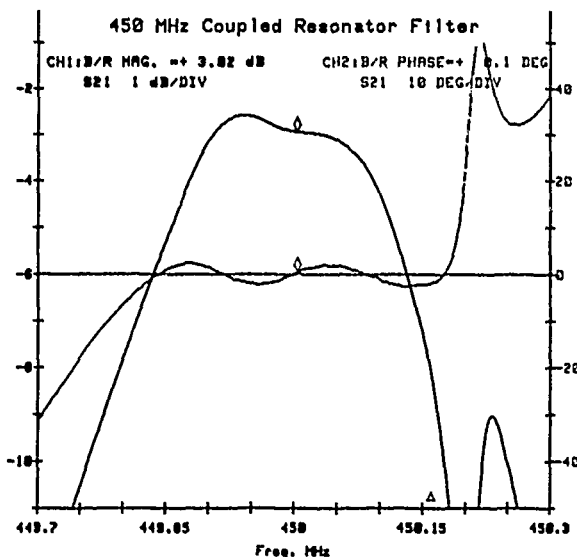


Figure 3 - Coupled Resonator Filter Response

Figure 3 gives the passband attenuation and linearised phase response of a typical coupled resonator filter at 450 MHz centre frequency. Main features are:

Insertion loss	2.9 dB
Bandwidth	300 kHz
Fractional bandwidth	0.07%
Phase ripple	<5° pk-pk

As resonator design is well documented (e.g. [9]) this low loss approach is ideal for narrow bandwidth front-end filtering at frequencies up to 1.2 GHz using resonators with $\lambda_0/4$ transducer geometries, where λ_0 is the acoustic wavelength at the device centre frequency.

Single Phase Unidirectional Transducer

The SPUDT approach requires only a single photolithographic step. The pattern resolution of $\lambda_0/4$, allows application to frequencies up to 1.2 GHz without the need for sophisticated manufacturing techniques. Other means of achieving unidirectionality were ruled out for the following reasons:

Manufacturing complexity caused by the need for several metal film depositions.

Fine line geometries, in some cases as small as $\lambda_0/20$ which limits upper frequency operation using standard photomask production methods.

Complexity of design in terms of both analysis and synthesis procedures

Complexity of application in terms of the number and tolerancing of components in the matching network.

The group type SPUDT comprises a ladder structure of several active rungs with a low loss reflector array distributed throughout the transducer in the spaces between the rungs (Fig. 4). The reflectors are offset $\pm \lambda_0/8$ from the centres of transduction, the direction of offset being dependent upon the substrate material [4,10]. On quartz the offset is away from the other transducer whilst on LiNbO_3 the offset is reversed.

GROUP TYPE SINGLE PHASE UNIDIRECTIONAL TRANSDUCERS (REF. LEWIS)

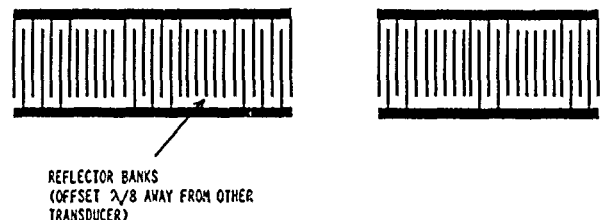


Figure 4 - Group Type SPUDT

Lewis [10] has shown that performance can be improved by using "blooming" anti-reflection strips either side of each active rung, offset $\lambda_0/4$ towards the rung. The rung to rung spaces are therefore filled with two types of reflectors. Figure 5 shows one repeat section of a SPUDT clearly indicating the relative offsets of the blooming and low loss reflectors and the $\lambda_0/4$ transducer geometries.

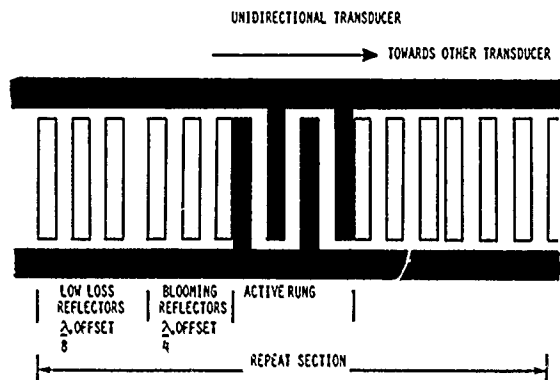


Figure 5 - SPUT - Repeat Section

The frequency response of this structure can be readily analysed to first order using a simple impulse response model. To demonstrate this device type, a low loss filter suitable for timing recovery application in a PCM transmission system has been designed and manufactured. Details of this filter are given in Table 1.

565 MHz Low Loss Filter

	Transducer No.1	Transducer No.2
Rungs	12	10
Finger pairs/rung	10	15
Blooming reflectors/rung	20	30
Low loss reflectors/rung	24	36

Table 1

The transducers are designed to give maximum possible interlacing of the individual sidelobes for the required bandwidth thus reducing these responses to a minimum.

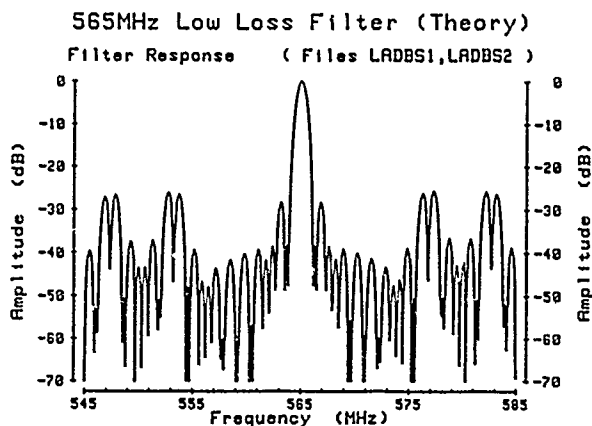


Figure 6 - Low Loss Filter - Theoretical Response

Results - Figure 6 gives the theoretical response of this device clearly showing the sidelobe roll-off characteristics of ladder structure transducers. The close-in sidelobes are at least 20 dB below the main response which is sufficient for many applications.

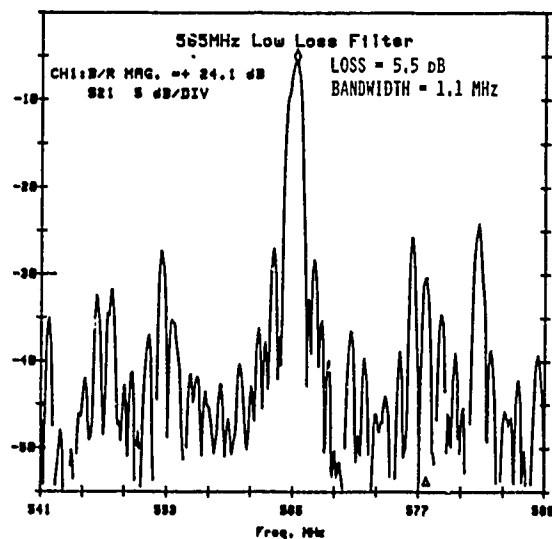


Figure 7 - Low Loss Filter - Real Response

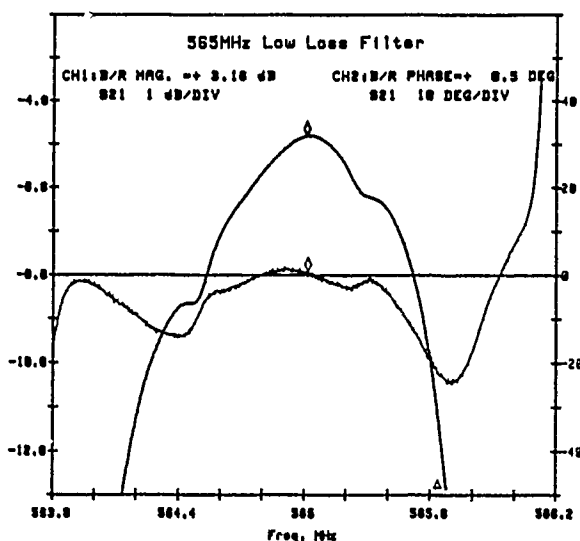


Figure 8 - Low Loss Filter - Passband

Figures 7 and 8 show the response of the real device housed in a T08 package in a tuned configuration with a series and shunt inductor per transducer. Main characteristics are:

Centre frequency	-	565 MHz
Insertion loss	-	5.5 dB
Bandwidth	-	1.1 MHz
Fractional bandwidth	-	0.2%
Sidelobe rejection	-	>20 dB
Linearised phase	-	<10° pk-pk

By comparison, a conventional unmatched bi-directional SAW filter designed for the same application has an insertion loss of 25 dB.

Maximally Distributed SPUDT

Some applications will require improved sidelobe roll-off, reduced phase ripple and smoother passband response. A second SPUDT design iteration has been completed where the number of transducer rungs is increased whilst the number of fingers per rung and rung to rung spacing is reduced. The number of finger pairs per rung is reduced to two with only four blooming and four low loss reflectors per repeat section. Full design details are given in Table 2.

450 MHz Maximally Distributed Low Loss Filter

	Transducer No.1	Transducer No.2
Rungs	25	23
Finger pairs/rung	2	2
Blooming reflectors/rung	4	4
Low loss reflectors/rung	4	4

Table 2

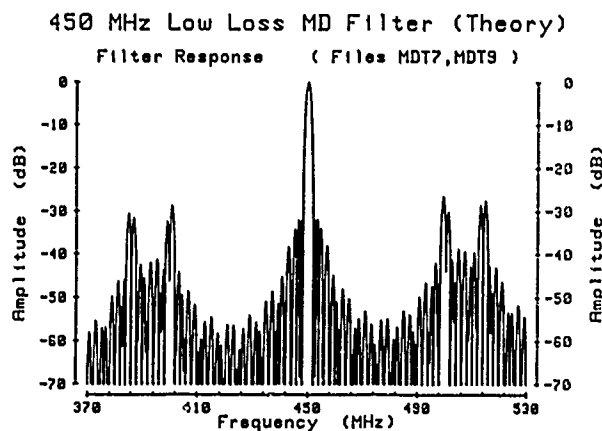


Figure 9 - Maximally Distributed SPUDT - Theoretical Response

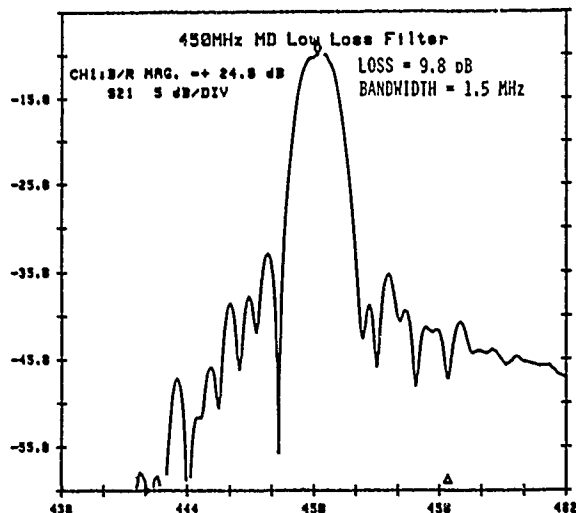


Figure 10 - Maximally Distributed SPUDT - Real Response

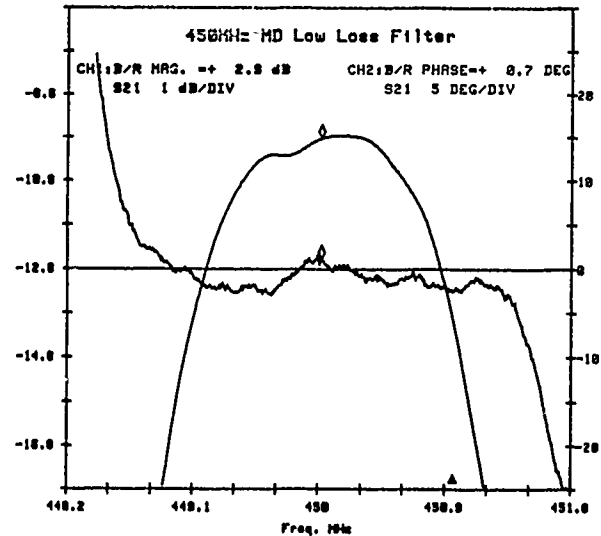


Figure 11 - Maximally Distributed SPUDT - Passband

Results - Figure 9 shows the predicted response indicating the improved sidelobe rejection. Figures 10 and 11 are plots of the real device in a matching configuration of series capacitor and shunt inductor per transducer. Main characteristics are:

Centre frequency	-	450 MHz
Insertion loss	-	9.8 dB
Bandwidth	-	1.5 MHz
Fractional bandwidth	-	0.33%
Sidelobe rejection	-	>25 dB
Linearised phase	-	<5° pk-pk

The design has improved passband ripple, phase linearity and wider bandwidth with the major trade-off being a greater insertion loss. For comparison, conventional unmatched filters have losses of 20 dB at these frequencies.

Advantages of SPUDT

Results detailed above show that the SPUDT and maximally distributed version have many advantages where narrow band filtering is required:

- Insertion loss as low as 5.5 dB
- Low amplitude ripple
- Low phase ripple
- Sidelobe rejection > 20 dB
- Ease of fabrication
- Simple analysis
- Simple application

TEMPERATURE STABILITY

The devices reported above were made on YX 39° quartz which has a similar temperature coefficient of delay (TCD) to ST-X quartz. Other orientations of quartz have been reported which have improved TCD. Two cuts of particular interest are SST [11] and LST [12]. SST quartz is doubly rotated and is therefore more difficult to align than ST-X. It also has improved TCD, a higher surface wave velocity and lower attenuation constant. The major drawback of SST is that in uncompensated designs, beam steering effects produce off-axis acoustic modes.

LST is a singly rotated cut which uses a leaky surface wave which has a lower coupling coefficient (K^2) than ST-X but a much higher acoustic wave

velocity. Shimizu [7] has shown that the TCD changes considerably for small variations in cut angle around the LST orientation which gives accuracy problems on crystal block alignment. However, the reported temperature stability shows a substantial improvement over ST-X quartz. Table 3 details appropriate ST-X, SST and LST parameters.

Quartz Temperature Characteristics		
Substrate	Designation YXL(θ)/ ψ	Frequency Stability $\Delta F/F$ (ppm) 100°C Range
ST-X	YXL(42.75°)	75
SST	YXL(49.2°)/22°	50
LST	YXL(-75°)	10-20

Table 3

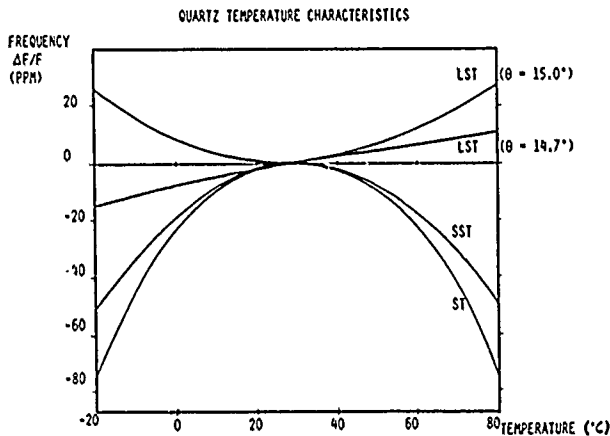


Figure 12 - Quartz Temperature Characteristics

Figure 12 compares these cuts relative to a common temperature turning point and shows the effect of a $\theta=0.3^\circ$ variation on the LST cut.

Results

To investigate the LST cut further standard designs have been fabricated on $YXL(-75^\circ)$ and have produced a measured TCD of

$$\approx 0.01 \times 10^{-9}/^\circ\text{C}^2$$

compared to the ST-X value of

$$\approx 0.03 \times 10^{-9}/^\circ\text{C}^2$$

Figure 13 gives the response of a 2-port SAW resonator on LST quartz showing an insertion loss of 15 dB with a Q of 4000 at 330 MHz. This design has not been optimised for the lower k^2 of LST and comprises 30 finger pairs per transducer and 500 fingers per mirror. The result illustrates the suitability of LST quartz for resonator designs and hence for low loss coupled resonator filters and stable SAW oscillators.

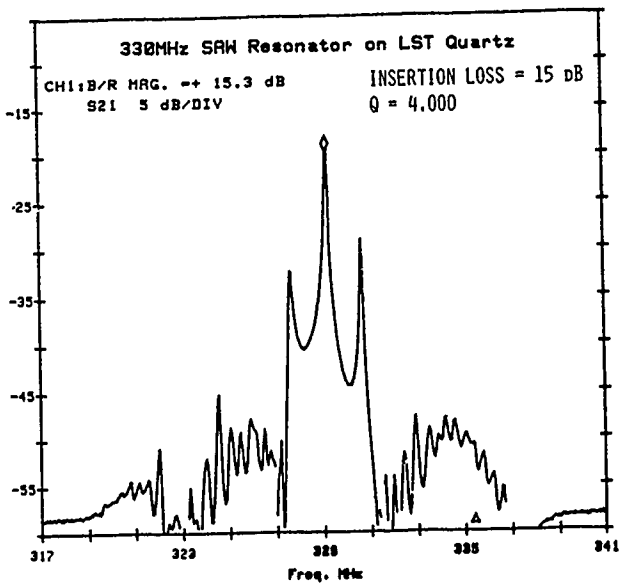


Figure 13 - Resonator on LST Quartz

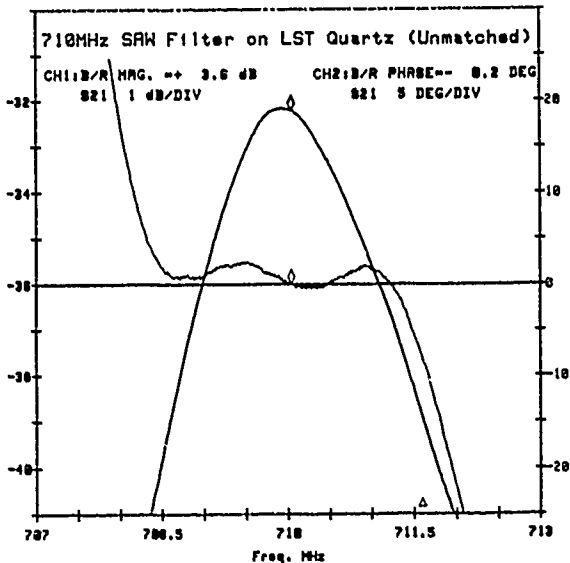


Figure 14 - Filter on LST Quartz

Figure 14 gives a plot of a conventional filter on LST quartz at 710 MHz. As expected the insertion loss of this device is higher than on ST-X, 32 dB compared to 27 dB, and the frequency is also higher, 710 MHz compared to 565 MHz.

These results suggest that LST quartz will support SPUDT designs correctly optimised for the lower k^2 value. Further work is now underway.

CONCLUSIONS

Coupled resonators are an attractive approach for narrow band filtering up to about 0.1% fractional bandwidth. Losses of 3 dB or less have been demonstrated on such structures.

It has been shown that the group-type SPUDT gives excellent results with insertion losses of 5.5 dB at a fractional bandwidth of 0.2% with good phase linearity and low passband ripple. The maximally distributed SPUDT gives increased bandwidth, improved phase linearity and smooth passband with the trade-

off of increased insertion loss. In both cases the out-of-band rejection shows sidelobes at least 20 dB below the main response.

The simplicity of fabrication will allow this technique to be applied to filters at frequencies up to 1.2 GHz using conventional photolithographic methods.

This structure lends itself to weighting techniques such as finger withdrawal, apodisation and block apodisation which will further improve device performance.

For improved temperature stability both resonator and conventional filter structures have been produced on LST quartz. Results suggest that low loss coupled resonator filters can be fabricated on LST quartz and work is now underway to optimise resonator and SPUOT designs for this new quartz cut.

ACKNOWLEDGEMENTS

The authors wish to thank Dr Meirion Lewis of RSRE, Malvern for discussions and advice, and also Mr Brian Duxon and Mr Mark Cracknell for manufacturing and tuning the SAW devices.

This work is supported by the United Kingdom Science and Engineering Research Council through the Interdisciplinary Higher Degrees Scheme of Aston University, Birmingham, UK.

REFERENCES

- [1] F G Marshall, E G S Paige, A S Young, "New Unidirectional Transducer and Broadband Reflector of Acoustic Surface Waves," IEE Electronic Letters, Vol 7, No 21, pp 638-640, October 1971.
- [2] R C Rosenfield, R B Brown, C S Hartmann, "Unidirectional Acoustic Wave Filters with 2 dB Insertion Loss," Proc IEEE Ultrasonics Symposium, 1974, pp 425-428.
- [3] K Yamanouchi, F M Nyffeler, K Shibayama, "Low Insertion Loss Acoustic Surface Wave Filter Using Group-Type Unidirectional Interdigital Transducer," Proc IEEE Ultrasonic Symposium, 1975, pp 317-321.
- [4] C S Hartmann, P V Wright, R J Kansy, E M Garber, "An Analysis of SAW IDTs With Internal Reflections and the Application to the Design of Single Phase Unidirectional Transducers," Proc IEEE Ultrasonics Symposium, 1982, pp 40-45.
- [5] M F Lewis, "SAW Filters Employing Interdigitated Interdigital Transducers," Proc IEEE Ultrasonics Symposium, 1982, pp 12-17.
- [6] M F Lewis, "Low Loss SAW Devices Employing Single Stage Fabrication," Proc IEEE Ultrasonics Symposium, 1983, pp 104-108.
- [7] Y Shimizu, M Tanaka, T Watanabe, "Leaky Surface Wave on Quartz and New Cut with Extremely Small Temperature Coefficient," Proc IEEE Ultrasonics Symposium, 1985.
- [8] P C Meyer, D Gunes, "Design and Fabrication of SAW Multipole Filters," Proc IEEE Ultrasonics Symposium, 1983, pp 66-71.
- [9] L A Coldren, R L Rosenberg, "Surface Acoustic Wave Resonator Filters," Proc IEEE, Vol 67, No 1, pp 147-158, January 1979.
- [10] M F Lewis, "A Study of Group-Type Single Phase Unidirectional SAW Transducers on LiNbO_3 and Quartz," RSRE, Malvern, UK, Memorandum No. 3833, June 1985.
- [11] T Lukaszek, A Ballato, "What SAW Can Learn from BAW: Implications for Future Frequency Control, Selection and Signal Processing," Proc IEEE Ultrasonics Symposium, 1980, pp 173-183.
- [12] Y Shimizu, M Tanaka, "New Cut of Quartz for SAW Devices with Extremely Small Temperature Coefficient," IEE Electronic Letters, Vol 21, No 6, pp 225-226, March 1985.

WIDEBAND TIMING TANK FILTERS FOR DIGITAL TRANSMISSION SYSTEMS

S. Yamamoto, T. Gounji and J. Shimizu

FUJITSU LIMITED

1015 Kamikodanaka, Nakahara-ku, Kawasaki 211, Japan

Abstract

LC filters have been used in the wideband timing tank filters of primary digital stage (1.544 MHz, 2.048 MHz) and secondary digital stage (6.312 MHz, 8.448 MHz) transmission systems, while helical filters have been used in the tertiary (44.736 MHz) digital stage. However, because of the recent developments in communication LSIs, smaller, more economical timing tank filters are required.

In response to these requirements, we have advanced the development of wideband timing tank filters using LiTaO₃ resonators, which have a large electro-mechanical coupling coefficient. For primary and secondary digital stage filters, thickness shear mode LiTaO₃ resonators, which have a zero first order temperature coefficient, were specially developed to achieve high stability. To achieve miniaturization, we adopted ladder-type filter construction, which needs no electrical matching circuits at the input or output and uses only two resonators and a condenser.

For the tertiary digital stage filters, BGS wave LiTaO₃ resonators utilizing edge reflection were specially developed to achieve easily-fabricated, miniature wideband filters.

These filters have a very small volume (1.4 cm³), about 1/8 that of LC filters and 1/30 that of helical filters. Their frequency characteristics fully satisfy the requirements, such as an insertion loss of less than 3 dB, a loaded Q of 50 to 120, an attenuation at half center frequency of more than 38 dB. Moreover, we carried out a variety of reliability tests, to investigate the effect of temperature, vibration, shock, aging etc., the good results indicating that practical use is possible. Nowadays these filters are being applied to bipolar/unipolar signal convert modules in a variety of digital transmission equipment, and good performance has been obtained.

Introduction

A timing tank filter is a device which regenerates the phase of transmitted pulse signals to assure correct transmission. The Figure 1 shows the usable range of a timing tank filter. There are two kinds of timing tank filter, that is, narrowband (loaded Q \approx 1000) and wideband (loaded Q \approx 100). The latter is mainly used for PCM primary, secondary and tertiary digital stages on the basis of digital hierarchy. To date, LC filters have been used for primary and secondary digital stages, and helical filters for the tertiary digital stage. However, due to the development of rapid communication LSIs, a need has arisen for a timing tank filter small enough to be used with LSI components. The filter should be highly stable and require no frequency adjustment circuits. The above filters cannot satisfy these requirements.

To solve this problem, we have developed LiTaO₃ resonators, which have a low capacitance ratio and a good temperature characteristic, and have investigated their application to timing tank filters. Consequently, we succeeded in developing miniature high-performance filters satisfying severe requirements and applied them successfully to transmission equipment. From now on, we will refer to a wideband filter using LiTaO₃ resonators as a LiTaO₃ filter.

Requirements

In order to be usable in equipment the LiTaO₃ filters must be of the specifications shown in Table 1.

Filter design

To achieve the desired miniaturization, we set out to devise a filter which does not require any electrical matching circuits at the input and output. Therefore, the filter had to be designed carefully so that both the input and output impedances met the specification. Two types of filter were possible candidates, the first being the monolithic filter, and the second, the electrically-coupled filter. The latter consists of two piezo-electric resonators and one condenser, and has the advantage of having an easily adjustable coupling degree, which can be made stronger than that of the monolithic filter. For these reasons, we decided to adopt the electrically-coupled filter.

An effective parameter theory was used in conjunction with a ladder circuit in the design of the filter. In general, a timing tank filter must have good characteristics such as a linear phase angle, a flat amplitude and a low passband loss. The designed filter must possess these characteristics and have the formerly described input and output impedances, that is less than 2 k Ω . We also investigated the Bessel and Butterworth characteristics. The electrically coupled filter was able to satisfy all the requirements, leading us to adopt it. Figure 2 shows the design procedure. Figure 2 (a) shows the Butterworth normalized lowpass filter, where $R_s = 1.0$ and $S_1 = S_2 = \sqrt{2}$. This filter is transformed to a band pass filter as shown in (b) and subsequently to a circuit by means of a gyrator transformation as shown in (c). However, a piezoelectric resonator has a parallel electrostatic capacitance C_0 (antiresonance) and a direct resonant resistance R , and these parameters have to be added to the circuit in (c). Because of this, we cannot determine the optimum design, other than by approximation. Figure 2 (d) shows the equivalent circuit of LiTaO₃ filter. Some stray capacitance must be considered in addition to the above. In particular, it is very important to compensate for stray capacitance C_s , parallel to coupling condenser C_k , because it has a great influence on the passband filter characteristics. Next, Figure 3 shows the equivalent circuit of the

Filter Item	Primary digital stage	Secondary digital stage	Tertiary digital stage
Center frequency f_0	1.544 MHz 2.048 MHz	6.312 MHz 8.448 MHz	44.736 MHz
Insertion loss	≤ 3 dB	≤ 3 dB	≤ 3 dB
Loaded Q	50 - 70	100 - 120	50 - 90
Phase stability (guaranteed for 20 years)	(1) 90 \pm 20 deg.	(1) 90 \pm 20 deg.	(2) -38 \pm 20deg.
Attenuation in stop band	≥ 36 dB 0.06 f_0 - 0.5 f_0 ≥ 22 dB 2 f_0 - 3 f_0	≥ 36 dB 0.06 f_0 - 0.5 f_0 ≥ 22 dB 2 f_0 - 3 f_0	≥ 36 dB at 0.5 f_0 ≥ 18 dB 2 M - 3 MHz
Spurious	≥ 10 dB at 3 f_0 , 5 f_0	≥ 10 dB at 3 f_0 , 5 f_0	≥ 36 dB at 0.5 f_0 ≥ 18 dB 50 M - 70 MHz
Input/Output impedance	≤ 2 k Ω	≤ 1 k Ω	$\leq 150 \Omega$
Volume	≤ 1.5 cm ³	≤ 1.5 cm ³	≤ 1.5 cm ³

- (1) Contains initial variation, temperature characteristic (0°C - 70°C) and aging characteristic.
(2) Contains temperature characteristic (0°C - 70°C) and aging characteristic.

Table 1 Performance required of LiTaO₃ filters

LiTaO₃ filter of the tertiary digital stage. The capacitance ratio of this filter's resonator is nearly twice that of the resonators of the LiTaO₃ filters in the primary and secondary digital stages. Therefore, a filter similar to that shown in Figure 1 (d) could not satisfy the stop band attenuation requirement (36 dB or more at half center frequency). Then the best method, enabling the realization of the above requirement's and miniaturization, was investigated. Consequently, a high-pass filter, composed of two chip coils (L_1 , L_2) and two chip condensers (C_1 , C_2), was added to the input side as shown in Figure 3. The optimum cut-off frequency which has no influence on the insertion loss and loaded Q of the filter and assures the necessary stopband attenuation proved to be 28 MHz. On the other hand, a great variation in the chip coil inductance disturbs the stability of the phase angle at the filter center frequency. Its temperature coefficient is not good, being a maximum of +5 %/°C. So, the temperature coefficient of the chip condenser necessary to compensate for these temperature characteristics was calculated, leading us to adopt a chip condenser whose temperature coefficient is -750 to -900 ppm/°C.

Resonators

A resonator with a large electro-mechanical coupling coefficient is necessary to the realization of a wide band timing tank filter (loaded Q = 100). Nowadays such resonators include ceramic, LiNbO₃ and LiTaO₃ resonators. The LiTaO₃ resonator has the best temperature characteristic of these. Therefore, the LiTaO₃ resonator was adopted.

(1) Specifications

A filter must not meet electrical matching circuits at the input or output, have low loss and high stability. Therefore, a resonator must satisfy the following specifications.

Resonator Item	For primary and secondary digital stages	For tertiary digital stage
Variation in inductance	$\leq \pm 10\%$	$\leq \pm 10\%$
Variation in resonant frequency	$\leq \pm 350$ ppm	$\leq \pm 450$ ppm
Resonant resistance	$\leq 100 \Omega$	$\leq 30 \Omega$
Frequency temperature characteristic (10°C - 70°C)	$\leq \pm 400$ ppm	-1500 ppm to +650 ppm

Table 2 Resonator specifications

(2) Vibration mode

(a) Filter for primary and secondary digital stages

In order to assure the frequency temperature characteristic, a thickness shear fundamental mode in a rotated X-cut LiTaO₃ plate², with a first order temperature coefficient of zero, was adopted. Also, an energy-trapped resonator and simple support method was adopted to obtain a high Q-factor. Figure 4 shows the orientation of the resonator.

To assure the resonator's inductance variation, the dimensions must be chosen so that only fine frequency adjustment is necessary. Figure 5 and Table 3 show an example resonator and measured characteristics, respectively.

(b) Tertiary digital stage filter

A bulk wave resonator has many problems of fabrication and reliability because of its extremely small size. We have advanced the development of the BGS wave^{3, 4} resonator using edge reflection on a rotated Y-cut LiTaO₃ plate.^{5, 6} This resonator has many advantages, such as easier fabrication than a bulk resonator, smaller size and lower capacitance ratio than a SAW resonator, not needing grating reflectors. However, the temperature characteristic of this resonator may be poor as its frequency temperature coefficient is about -30 ppm/°C. This problem was solved by reducing the value of the loaded Q as much as possible (loaded Q = 60). Figure 6 shows the orientation of the resonator. The propagation direction of the BGS wave corresponds to the X'direction in Figure 6. Figure 7 and Table 4 show the dimensions of the resonator and its measured characteristics respectively.

Filter construction

Bare chip type resonators were adopted as their size allows the realization of a miniature filter. The substrate is Al₂O₃ ceramic, this being very strong in thermal shock and its thermal expansion coefficient (6.7 ppm/°C) is close to that of the LiTaO₃ resonator (10 ppm/°C in the longitudinal direction). Ad-Pa thick film was adopted for the electrode patterns because of its high reliability. A thin conductive plate, which has a small slit, relaxes some stress at the center. It can relieve the thermal stress resulting from the difference between the thermal expansional coefficient of the LiTaO₃ resonator and that of the ceramic substrate. Figure 8 shows the variation in resonant frequency after heat cycle tests performed to analyze the effect of using the supporter. With the supporter, the resonant frequency variation is about half that obtained when the resonator is mounted directly on the substrate. Despite the above the BGS wave LiTaO₃ resonator is directly bounded to the substrate because we can assume that the influence of the stress is very small.

The filter is sealed in nitrogen gas within a 14-pin IC DIP package. Figures 9 and 10 are photographs of the primary digital stage LiTaO₃ filter (1.544 MHz) and the tertiary digital stage (44.736 MHz) respectively.

Experimental results

(1) LiTaO₃ filters for primary and secondary digital stage

Figure 11 shows the experimental filter and the LC filter together, the dimensions are 20.4L x 12.8 W x 5.5 H (mm) and the volume (1.4 cm³) is only 1/8 that of an LC filter.

As a representative example, the LiTaO₃ filter whose center frequency (f_0) is 1.544 MHz is described. Figure 12 shows the equivalent circuit of the experimental filter. The input and output impedances are both 1.475 k Ω . Figure 13 and Figure 14 show the phase and attenuation characteristics in the vicinity of the passband, respectively. The insertion loss and phase angle at the center frequency (f_0) are 0.4 dB and +96 deg., and the loaded Q and the phase angle inclination are 59 and 6.50 deg./kHz. Figure 15 shows the stopband attenuation characteristic. The spurious characteristic is quite

good except for the third overtone response. And the attenuations in the frequency ranges from 0.06 f_0 to 0.5 f_0 and 2 f_0 to 3 f_0 are 38 dB, or more and 26 dB, or more respectively, thus satisfying the formerly described requirements. Figure 16 shows the temperature characteristics of the experimental filters, where the reference temperature is 25°C. The variation of phase and insertion loss in a wide temperature range from 0°C to 70°C are -3.5° to +2° and -0.25 dB to +0.20 dB. The phase variation of this filter is nearly 1/4 that of an LC filter, indicating high stability. This meets the requirements.

(2) LiTaO₃ filter for tertiary digital stage

Figure 17 shows the experimental filter and helical filter together. Its dimensions are as 20.4 L x 12.8 W x 5.5 H (mm) and its volume (1.4 cm³) is 1/30 that of helical filter. Figure 18 shows the equivalent circuit of the experimental filter whose center frequency (f_0) is 44.736 MHz. The input and output impedances are both 105 Ω . Figures 19 and 20 show the phase and attenuation characteristics in the vicinity of the passband. The insertion loss and phase angle at the center frequency (f_0) are 1.4 dB and -38.5°, and the loaded Q and phase angle inclination are 63 and 0.217 deg/kHz. The insertion loss of this filter is only 1/2 that of a helical filter. Figure 21 shows the stopband attenuation characteristic. The attenuations at half center frequency (± 22 MHz) and in a frequency range from 50 MHz to 70 MHz are 43 dB and 20 dB, thus satisfying the formerly described requirements. Figure 22 shows the temperature characteristics of the experimental filters, where the reference temperature is 25°C. The variations of phase angle and insertion loss in a wide temperature range from 0°C to 70°C are -7 deg. to +14 deg. and -0.3 dB to +0.4 dB, which are as good as that of helical filter. These characteristics fully satisfy the requirements.

Reliability test results

The authors have carried out a variety of reliability tests to test the effects of vibration, shock, heat cycle, dry heat and aging and so on. The test conditions were shown in Table 5.

(1) LiTaO₃ filter for primary and secondary digital stages

As an example, the test results for a resonator of a 1.544 MHz LiTaO₃ filter are described. Figure 23 shows the vibration test result. The variation of the resonant frequency is -16 ppm minimum and +25 ppm maximum (specification: +150 ppm). Figure 24 shows the shock test result. The variation in the resonant frequency is +15 ppm minimum and +50 ppm maximum (specification: +150 ppm). Figure 25 shows the heat cycle test result.

The variation of the resonant frequency is -27 ppm minimum and +50 ppm maximum (specification +150 ppm). Figure 26 shows the result of the dry heat test. The variation of the resonant frequency is -42 ppm maximum and +23 ppm minimum satisfies (specification: +150 ppm). Figure 27 shows the aging characteristic at room temperature. The aging coefficient of the resonant frequency is -35 ppm/decade minimum satisfies (specification: -100 ppm/decade to +40 ppm/decade). Figure 28 shows the aging characteristic at 80°C. In this case, the samples are resonator for 6.312 MHz LiTaO₃ filter. The measured result satisfies the above specifications.

The reliability test results showed that the specifications were met in all respects.

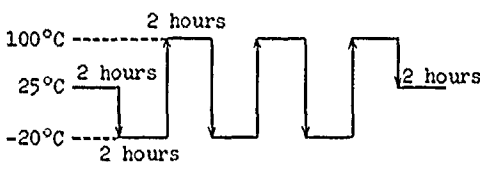
Item	Condition
Vibration	Frequency: 10 Hz to 55 Hz Amplitude: 1.5 mm Period: 1 minute Direction: X, Y, Z Time: 2 hours each direction
Shock	Acceleration: 50 g Time: 11 m seconds Direction: +X, +Y, +Z 3 times each direction
Heat cycle	Number of cycles 3 cycles 
Dry heat	Temperature: 100°C Duration times: 100 hours
Aging	At room temperature (25 \pm 3°C) or 80°C \pm 1°C

Table 5 Reliability test conditions

X, Y, Z directions correspond to the length, width and thickness of a filter respectively. The measurements were carried out at room temperature.

(2) LiTaO₃ filter for tertiary digital stage.

The test results of a LiTaO₃ filter with a center frequency is 44.736 MHz are described. Figure 29 shows the vibration test result. The variation of the phase angle at the center frequency is -0.5 deg. at the minimum and +0.4 deg. at the maximum, (specification: \pm 1 deg.). Figure 30 shows the shock test result. The variation of the phase angle at the center frequency is -0.5 deg. at the minimum and +0.2 deg. at the maximum (specification: \pm 1 deg.)

Figure 31 shows the heat cycle test result. The variation of the phase angle at the center frequency is -0.02° minimum and +1.4° maximum (specification: \pm 2 deg.)

Figure 32 shows the aging characteristic at 80°C. The aging coefficient of the phase variation at the center frequency is -1.0 deg/decade minimum (specification: -1.5 deg./decade to +0.5 deg./decade).

Application of Digital Equipment

The developed LiTaO₃ filters are contributing to the realization of small, high performance low cost transmission equipment.

LiTaO₃ filters for primary and secondary digital stages are being applied to Bipolar/Unipolar and Unipolar/Bipolar converting modules used in multiplex transmission equipment where they have performed well. Figure 33 is a photograph of a 1.544 M b/s B/U converting module. LiTaO₃ filters are being applied to the tertiary digital stage of the B/U and U/B converting module used in high speed optical transmission equipment. Again, good results have been obtained. Figure 34 shows the 44.736 M b/s B/U and U/B converting module.

Conclusion

New wideband timing tank filters were realized by the use of high coupling LiTaO₃ resonators. These filters could be applied to digital transmission equipment. Moreover, these filters proved to be highly reliable. The features of the LiTaO₃ filters and their effect on performance are as follows.

- (1) The filters are small, being only 1/8 the size of an LC filter and 1/30 that of a helical filter. No electrical matching circuits are necessary at the input and output. Therefore, an LiTaO₃ filter is compatible with LSIs in both shape and size.
- (2) The insertion loss is less than 3 dB and the stopband attenuation is more than 38 dB in a frequency range from 0.06 f_0 to 0.5 f_0 , where f_0 is the center frequency. Therefore, an LiTaO₃ filter assures a high signal to noise ratio.
- (3) The temperature characteristics are excellent; nearly 1/4 that of an LC filter. Therefore an LiTaO₃ filter has a highly stable output waveform with only a little phase variation.
- (4) An external frequency adjustment circuit is not necessary. Therefore the cost is reduced.

Acknowledgments

The authors wish to thank Mr. Takamoto Kojima, Director of the Communication Systems Group, Mr. Kenji Shirai, Director of the Technology Development Dept. and Mr. Hideki Tomimaga, Director of the Circuit Device Dept. for their useful advice and continuous encouragement.

References

- (1) Kawana and H. Kawahata, "Preshift Mechanical Filter for Voice Frequency Telegraph Transmission System," IEEE 1979 Ultrasonic Symposium, P119, 1979.
- (2) Y. Fujiwara, S. Yamada and N. Wakatsuki, "Miniaturized LiTaO₃ Strip Resonator," Proc, 37th Annual Symposium on Frequency Control, P.343, 1983.
- (3) J. L. Bleustein, "A New Surface Wave in Piezoelectric Material", Appl. Phys, Letts., 13, P.412, 1968.
- (4) Yu. V. Gulyaev, "Electroacoustic Surface Waves in Solid," Sov. Phys. JETP Letters, 9, P37, 1969.
- (5) K. Nakamura, M. Kazumi and H. Shimizu, "SH-Type and Rayleigh-Type Surface Waves in Rotated Y-Cut LiTaO₃," Paper, Technical Group on Ultrasonics, IECE Japan, US77-42, P.31, 1977.
- (6) M. Ono, Y. Fujiwara and N. Wakatsuki, "Surface Shear Wave Resonators for VHF Range Utilizing Reflection at Edge Surfaces," IEEE 1984 Ultrasonic Symposium, P.218, 1984.

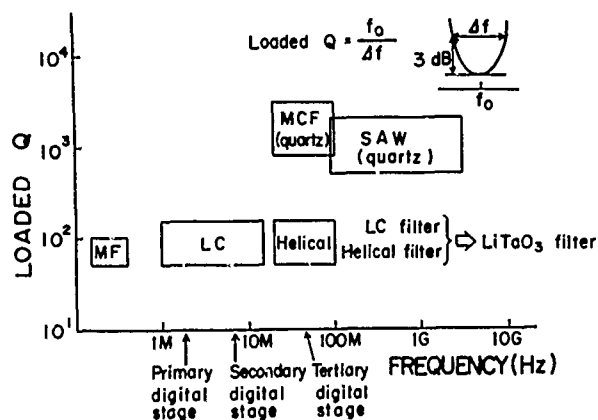


Figure 1 Usable range of timing tank filters

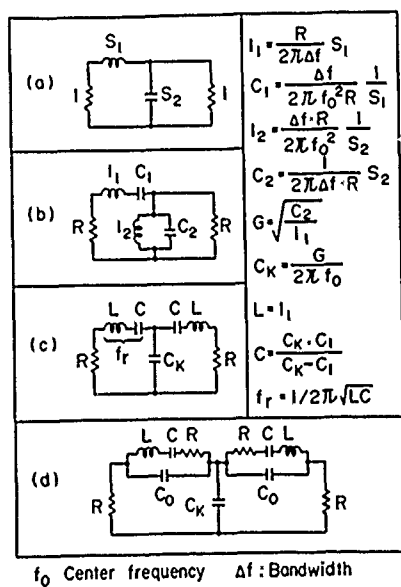


Figure 2 Filter design procedure

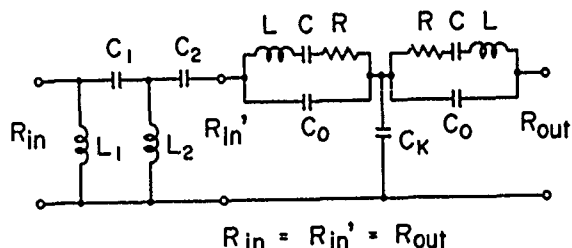


Figure 3 Equivalent circuit of LiTaO₃ filter of tertiary digital stage

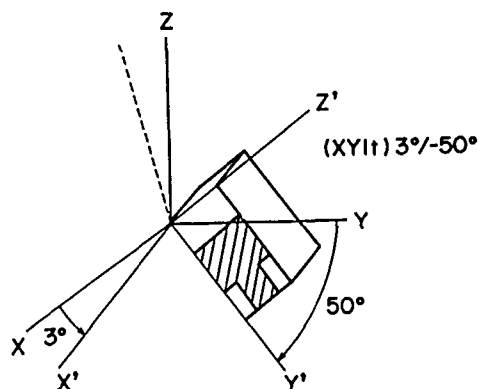


Figure 4 Rotated X-cut LiTaO₃ plate

Resonant frequency $f_r = 1.534044$ MHz

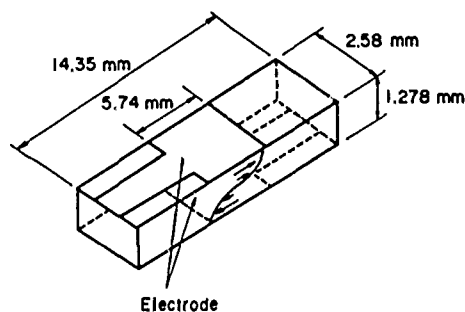


Figure 5 Schematic of Thickness shear mode LiTaO₃ resonator

Item	Value
f_r (MHz)	1,534044
L (mH)	12,0
C (pF)	0,8969817
R (Ω)	45,3
C_0 (pF)	6,7
Q_M	2553
γ	6,8
Second order temperature coefficient (ppm / °C ²)	0,050

Table 3 Equivalent constants of thickness shear mode LiTaO₃ resonator

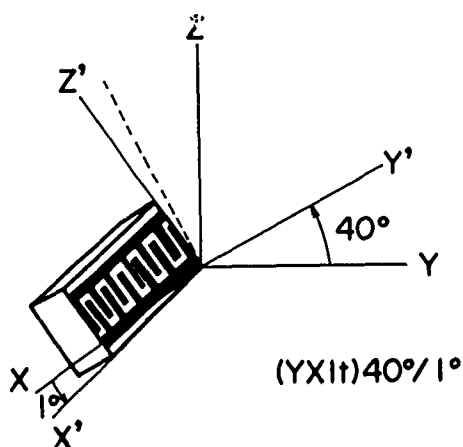


Figure 6 Rotated Y-cut LiTaO₃ plate

Resonant frequency $f_r = 44.6406$ MHz

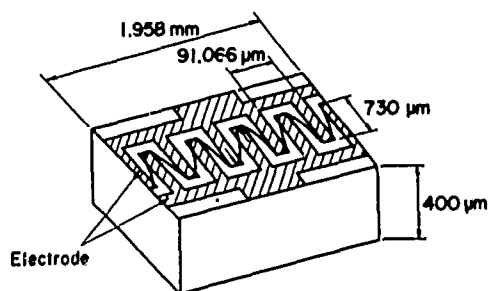


Figure 7 Schematic of BGS wave LiTaO₃ resonator

Item	Averaged value
f_r (MHz)	44.6671
L (μH)	31.7
C (pF)	0.400978
R (Ω)	19.9
C_0 (pF)	7.3
Q_M	490
γ	18.2
Temperature coefficient $\left(\frac{\text{ppm}}{^\circ\text{C}}\right)$	-29.7

Number of samples : 12

Table 4 Equivalent constants of BGS wave LiTaO₃ resonator

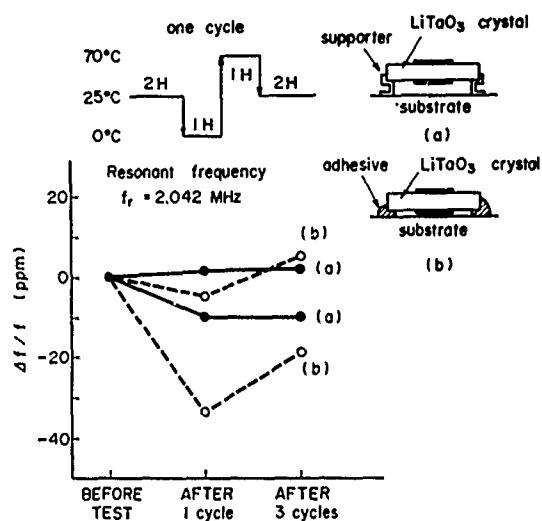


Figure 8 Effect of support method in heat cycle test

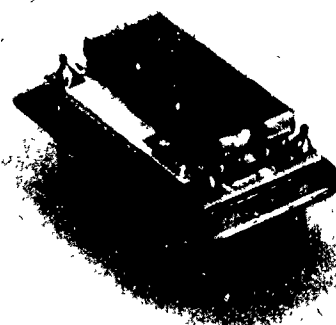


Figure 9 1.544 MHz LiTaO₃ filter of primary digital stage

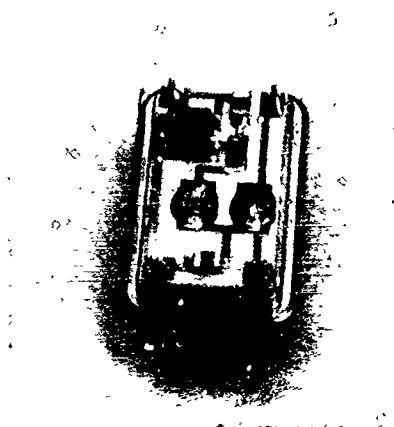


Figure 10 44.736 MHz LiTaO₃ filter of tertiary digital stage

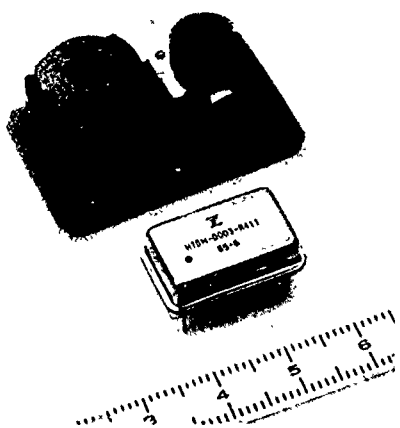


Figure 11 Size comparison between LiTaO_3 filter and LC filter

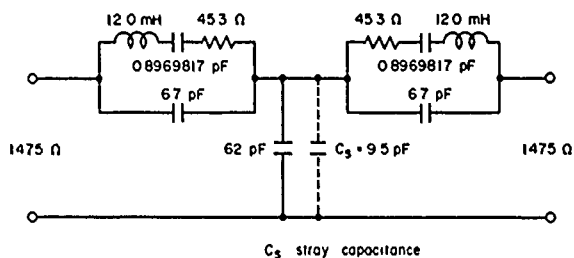


Figure 12 Equivalent circuit of experimental filter (1.544 MHz)

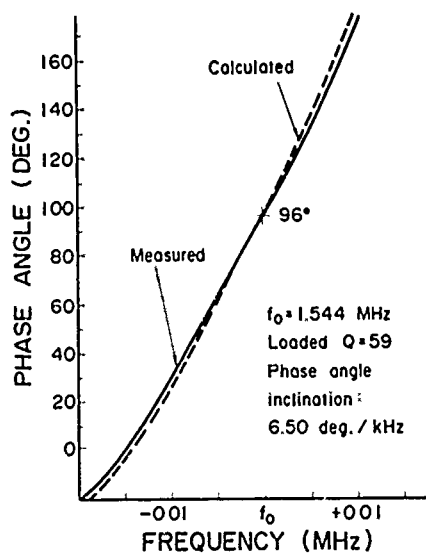


Figure 13 Passband phase characteristic

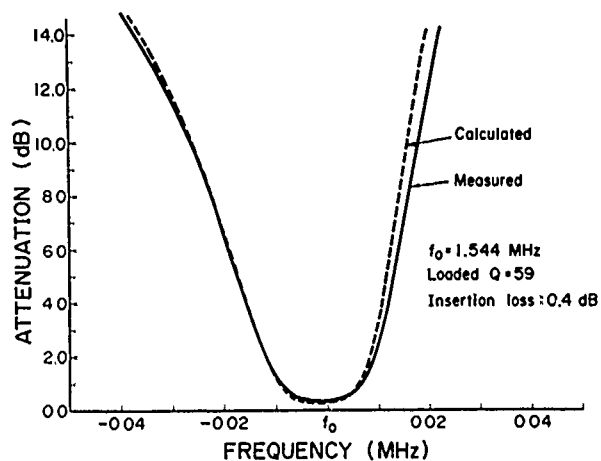


Figure 14 Passband attenuation characteristic

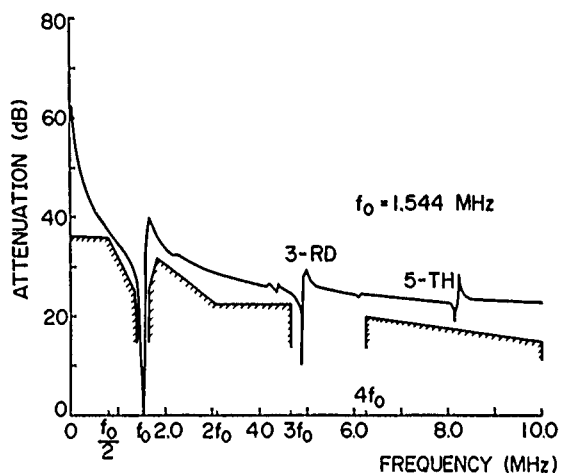


Figure 15 Stopband attenuation characteristic

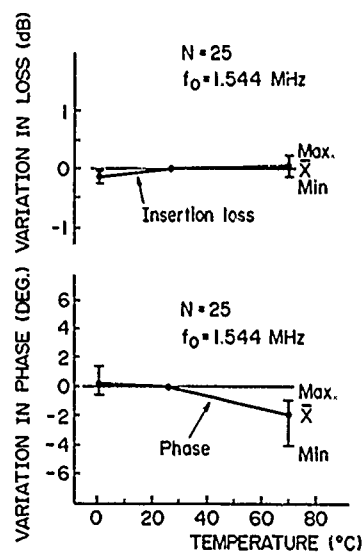


Figure 16 Temperature characteristic

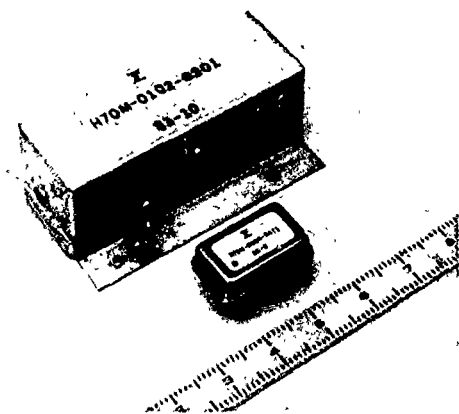


Figure 17 Size comparison between LiTaO₃ filter and helical filter

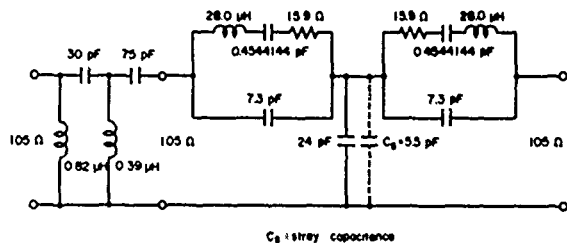


Figure 18 Equivalent circuit of experimental filter (44.736 MHz)

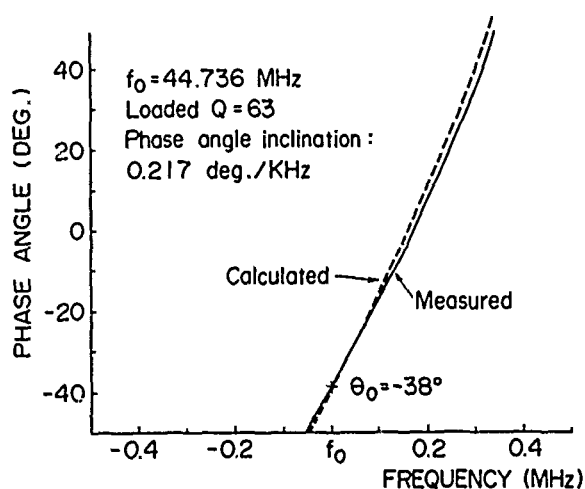


Figure 19 Passband phase characteristic

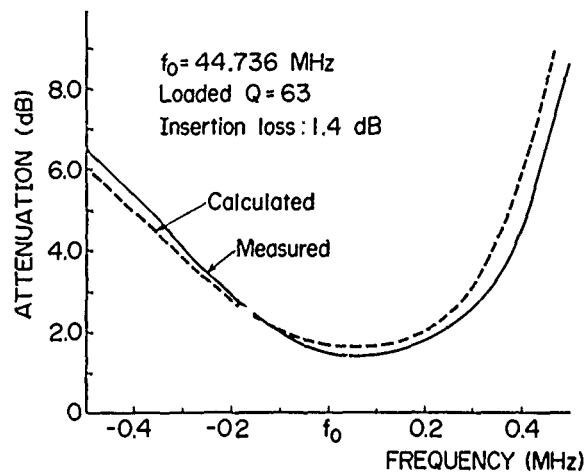


Figure 20 Passband attenuation characteristic

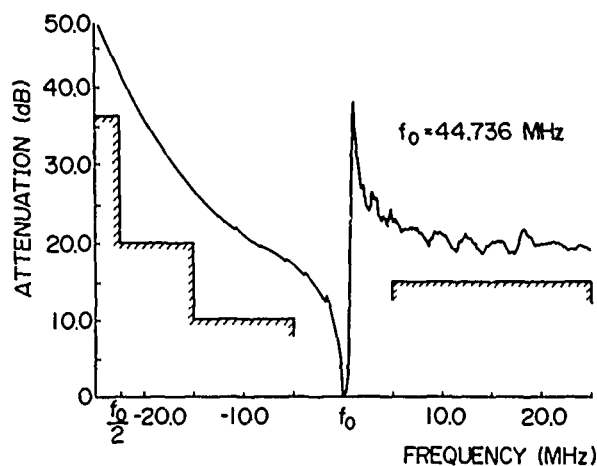


Figure 21 Stopband attenuation characteristic

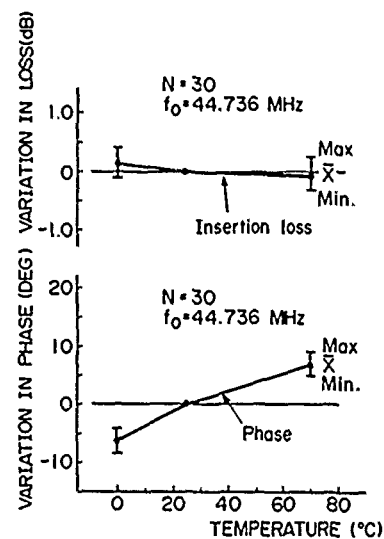


Figure 22 Temperature characteristic

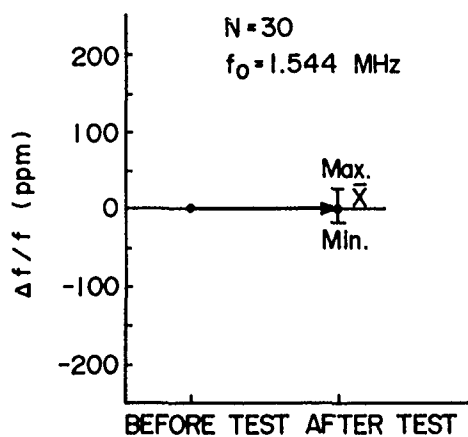


Figure 23 Vibration test result

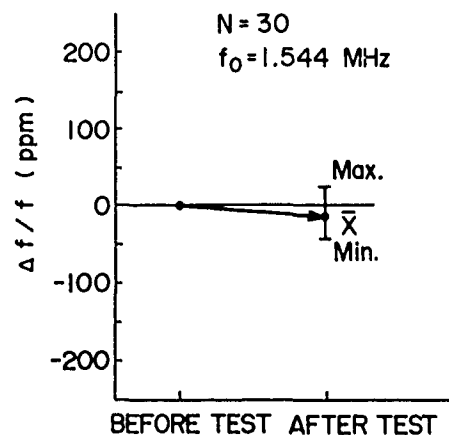


Figure 26 Dry heat test result

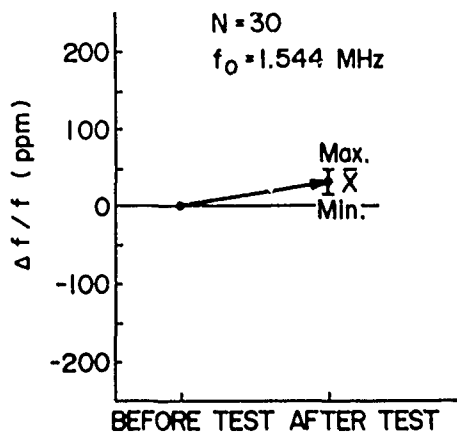


Figure 24 Shock test result

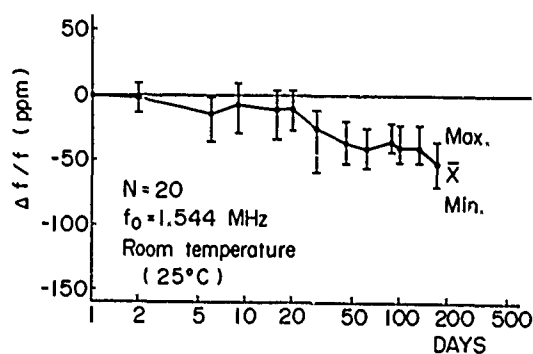


Figure 27 Aging characteristic at room temperature

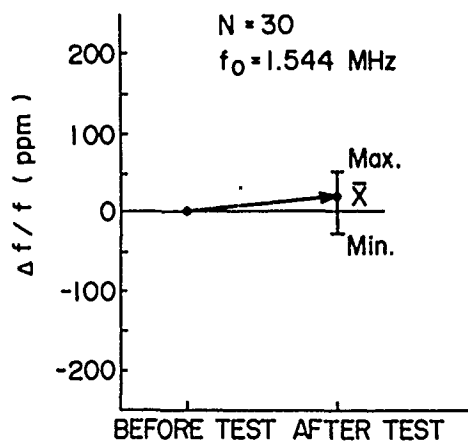


Figure 25 Heat cycle test result

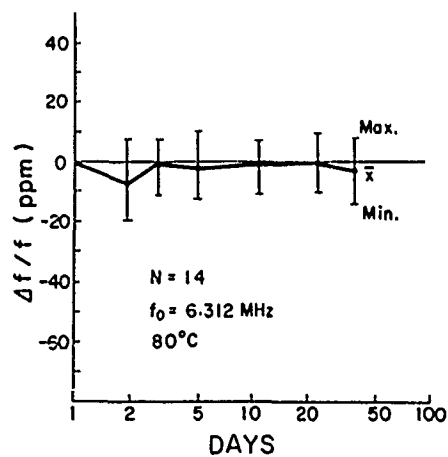


Figure 28 Aging characteristic at 80°C

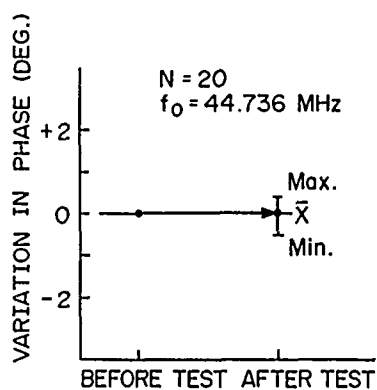


Figure 29 Vibration test result

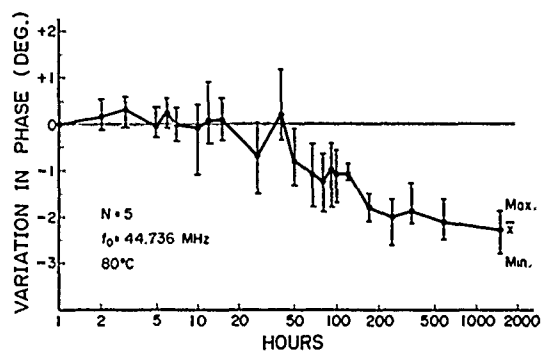


Figure 32 Aging characteristic at 80°C

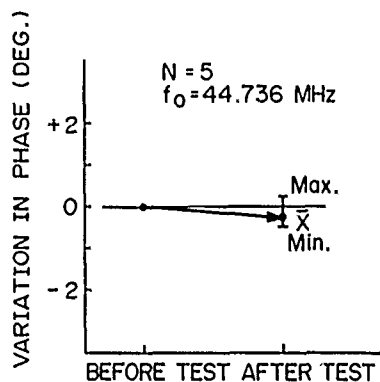


Figure 30 Shock test result

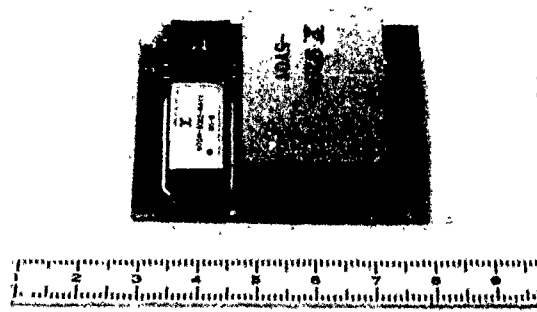


Figure 33 1.544 M b/s B/U converting module

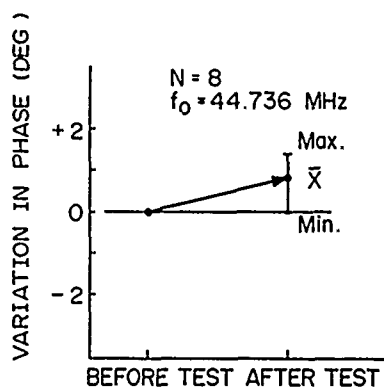


Figure 31 Heat cycle test result

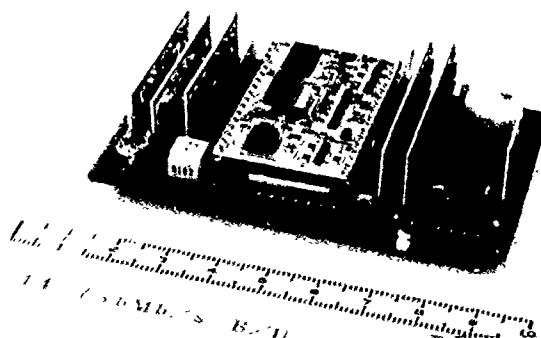


Figure 34 44.736 M b/s B/U-U/B converting module

ACOUSTIC CHARGE TRANSPORT
PRINCIPLES AND PERFORMANCE

M.J. Hoskins, M.J. Brophy, J.M. Dallesasse,
M.J. Miller, J.W. Peterson

Electronic Decisions Incorporated
Urbana, Illinois 61801

Summary

The basic operational principles and signal processing capabilities of the acoustic charge transport (ACT) device are presented. The ACT processor is a new type of high speed GaAs charge transfer device in which electron transport is accomplished using the traveling wave electric field of a surface acoustic wave (SAW). The inherently smooth and continuous nature of traveling wave transport yields very high transfer efficiency, large signal bandwidths and interference-free charge detection. Experimental results are presented for an ACT delay line operating at a 360 MHz SAW frequency with a transfer efficiency in excess of 0.99994 and a signal bandwidth of 180 MHz.

The nondestructive charge sensing (NDS) principles which form the basis for multiple-tapped delay line applications are discussed and experimental results are presented for ACT transversal filters incorporating NDS electrode arrays.

I. Introduction

The development of analog delay elements for integrated circuits receives increasing interest as GaAs I.C. technology matures at a rapid rate. This interest is motivated by the wide variety of analog signal processing functions which can be achieved by combining transversal filter devices with gain and control interface circuitry. In response to this, both the GaAs surface acoustic wave¹ and charge coupled device² (CCD) technologies have been studied.

In the last few years, a new GaAs device approach³⁻⁶ which combines some of the characteristics of both the SAW and CCD technologies was shown to have the capability to achieve high speed delay line and transversal filter functions. In the Acoustic Charge Transport (ACT) device, electron transport at the sound velocity is accomplished using the traveling wave electric field of a single frequency SAW. By analogy with a conventional CCD, the SAW functions as a built in "clock signal" to power the charge transfer of high mobility electrons in a buried channel structure. Experimental and theoretical⁷ work has shown that the inherently smooth and continuous nature of traveling wave transport in the ACT device yields high charge transfer efficiency, low interference charge detection, and large signal bandwidths while the

large electron mobility of GaAs permits high speed operation.

Recent advances in ACT device architecture which utilize proton bombardment for channel definition and isolation have permitted a large degree of flexibility for signal processor design while retaining excellent charge transfer performance. In this paper, the basic operational principles of the ACT device are summarized and the experimental performance of an ACT delay line and transversal filter based on the proton isolated architecture are presented. To the authors' knowledge, the charge transfer efficiencies obtained in these devices are the highest ever reported for any GaAs charge transfer device.

II. ACT Principles

A. Charge Transport

Figure 1a illustrates the side view structure of the basic ACT delay line used in early concept demonstration investigations⁴. An epitaxial GaAs n-p layer configuration is grown on a semi-insulating substrate such that the n-type layer thickness is roughly one half to one acoustic wavelength thick. Similar structures without the buried p-layer have also been used⁶. A Schottky barrier channel plate on the surface of the n-type layer defines the length of the ACT channel. Ohmic contacts at the input and output ends of the channel form lateral Schottky barrier diodes with respect to the channel plate permitting electron injection and extraction during operation. An interdigital transducer generates large amplitude surface acoustic waves which illuminate the transport channel structure. Although the piezoelectric coupling coefficient of GaAs is relatively small, efficient transduction of the SAW at a single frequency can be obtained by constructing the transducer with a long interaction region.

The buried transport channel is defined by applying biases to the structure which deplete the n-type layer from the top and the bottom via the Schottky channel plate and buried p-layer respectively. The plane at which these depletion regions just overlap represent a plane of minimum electron potential which tends to confine injected electrons near the center of the n-type active layer as shown in Fig. 1b. The piezoelectric traveling wave potential as-

sociated with the SAW is fully screened at the surface by the metallic channel plate and in the buried p-layer but the SAW potential magnitude peaks up near the center of the n-layer due to the rapid exponential decay of the screening functions. The SAW potential in the channel may be expressed as the product of an x directed traveling wave function and an "envelope" function which is only dependent on the space coordinate y.

$$\phi(x, y, t) = \phi_w(y) e^{j(kx - \omega t)}$$

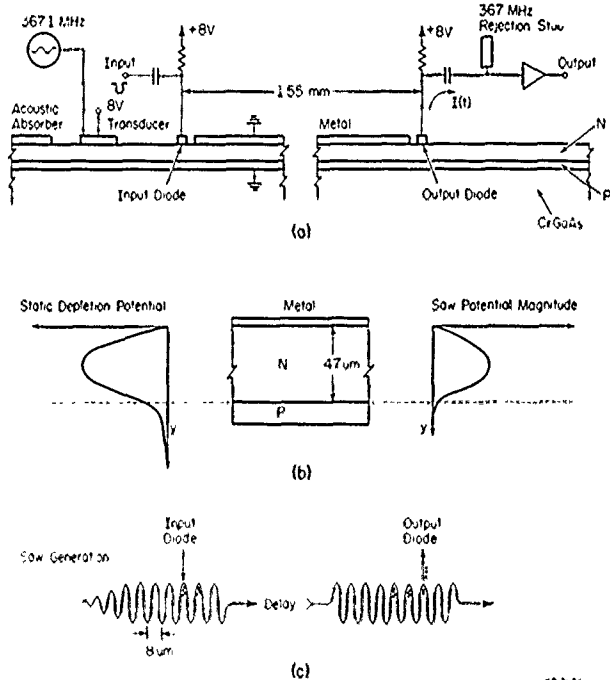


Fig 1. ACT device operation.

- a) Side view of the ACT structure
- b) Transport channel potentials
- c) Charge injection and detection

The envelope function $\phi_w(y)$, represents the magnitude of the SAW potential at any depth y. Both the depletion potential and the SAW potential provide electron confinement near the center of the n-layer, although the depletion potential is more dominant in this regard because it typically has a steeper gradient. Although it is not shown in Fig. 1a, separate bias voltages may be applied to the Schottky barrier and buried p-layer to alter the static depletion potential minimum point, and hence, the depth of transport. The ability to control the location of charge transport in the channel layer via bias variations is in part responsible for the excellent transfer efficiency obtained in the ACT device.

A depletion bias is also applied to the Schottky barrier formed by the SAW transducer to avoid screening of the transducer fields by the conductive n-layer. A properly constructed transducer is capable of generating

a SAW potential magnitude of a few volts which is more than sufficient to provide adequate charge transport capacity.

The superposition of the SAW and depletion potentials creates a series of two dimensional potential wells which propagate through the transport channel at the SAW velocity (2864 m/s). Electrons are injected into these potential wells by applying a negative going pulse to the input contact as shown in Fig. 1a. This pulse momentarily collapses the depletion region at the input end of the channel and the injected electrons are quickly bunched by the traveling wave potential into concise packets which are trapped by the SAW and constrained to move precisely at the sound velocity. After transport through the length of the channel, the electron packets are rapidly swept out of the wave by the large electric fields associated with the reverse biased lateral output diode resulting in a current spike in the output detection circuit. The charge injection, transport and detection operations are depicted one dimensionally in Fig. 1c. In analog delay line and transversal filter applications, a continuous quiescent electron current is established in the device to permit the application of both polarities of input signal swings.

A theoretical analysis was performed⁷ which reveals the fundamental characteristics of the ACT process. In particular, it is found that the transport properties may be parameterized in terms of a quantity defined as the channel constant β/ω where

$$\frac{\beta}{\omega} = \frac{q N_d}{\phi_0 \epsilon k^2}$$

Here, N_d is the doping density in the n-type channel layer, ϕ_0 is the SAW potential magnitude at the depletion potential minimum point, k is the acoustic wavenumber, q is the electronic charge and ϵ is the dielectric constant. The quantity β represents the depletion charge density while the quantity ω is the peak piezoelectric polarization charge density associated with the SAW. An interesting result from the transport analysis is that the ACT packet shape is primarily dependent on the magnitude of the channel constant. This fact is illustrated in Fig. 2 where the charge packet boundary shape is shown for various channel constants. For a channel constant near unity, the charge packet shape is nearly circular, while larger channel constants (corresponding to larger depletion fields) give approximately elliptic boundaries.

When the channel constant becomes very large, the charge packets are compressed into a sheet charge geometry due to the strong confinement effect of the depletion field. The packet shape is also slightly dependent on the charge load; the curves of Fig. 2 correspond to a fractional charge load of 1/2. As the charge load varies, the basic packet shape remains roughly the same, but the overall packet dimensions scale proportionately in both the x and y directions. At full charge load, the maximum packet length in the propagation direction (x) is approximately one half of an acoustic wavelength λ . By way

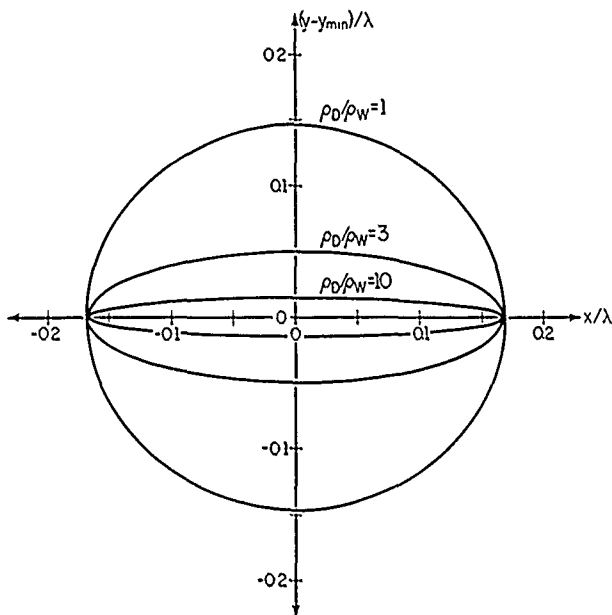


Fig 2. Charge packet boundary shape.
($\phi_{cl}/\phi_0=1/2$)

of example, in an ACT device operating at 360 MHz SAW frequency with a wavelength of 8 μm , the full charge load packet length is roughly 4 μm . If the channel constant is 3, the full charge load packet width is roughly 1 μm .

As the charge load increases, the SAW potential barriers which confine the packet charge are continually reduced until they completely vanish at full charge load. Hence, at very large charge loads, diffusion losses can disperse packet charge into neighboring potential wells creating a source of transfer inefficiency. Transfer inefficiency in a conventional CCD is defined as the fractional charge which is lost from the packet in each transfer. A similar definition applies to the ACT device if a transfer is considered to be the motion of charge over the distance of one SAW wavelength. The room temperature diffusion induced transfer inefficiency is plotted as a function of charge capacity with SAW potential magnitude as a parameter in Fig. 3 using the transport theory described in Ref. 7. The calculations are performed for a channel constant of 3. The charge capacity N represents the number of electrons per packet per centimeter of channel width, and it is independent of the SAW frequency. The charge capacity range over which the transport is significantly degraded by diffusion induced transfer inefficiency is a small fraction of the total dynamic range of the device for SAW potentials of 1V or more. For example if $\phi_0=1\text{V}$, and a transfer inefficiency of 10^{-5} is arbitrarily chosen as the minimum level of acceptable performance, then degradation of performance only occurs over the last 2.5 dB of full well signal charge. A good rule of thumb is that the first 10 KT/q of wave potential is required to produce a diffusion inhibiting barrier; the remainder of the wave potential contributes to charge

capacity. Hence, a simple analytic approximation to the charge capacity curves of Fig. 3 is

$$N(\text{e}^-/\text{cm}) \approx 53 \times 10^6 (\phi_0 - 0.26)$$

A 1 V wave can carry roughly 40 million electrons per packet per centimeter of channel width. The charge capacities predicted by the transport theory have been achieved in experimental devices. High transfer efficiencies have been obtained in experimental devices as will be shown later in this paper, but they tend to be limited by material impurity charge trapping effects rather than diffusion.

Although the ACT device operation appears similar to that of a conventional CCD, there are some significant differences. The transport of charge occurs continuously and entirely under the influence of constant electric drift field for all space and time. The charge packet shape and charge density are essentially invariant in the constant velocity traveling wave reference frame. The charge packets follow a straight line path at constant depth beneath the surface throughout the entire transport interval because the surface metal provides a very uniform surface potential boundary condition. All of these factors enhance the fundamental transfer efficiency capability of the device. Another interesting characteristic results from the automatic input sampling property of the device. As a SAW potential well propagates under the input contact, the charge potential equilibration of the packet with the input contact is naturally cut off by the trailing barrier of the SAW potential. Hence, input signals can be applied directly to the device asynchronously without clocked gating circuits. Finally, the single frequency SAW "clock" signal minimizes signal interference in the charge detection circuits and permits efficient asynchronous non-destructive charge detection.

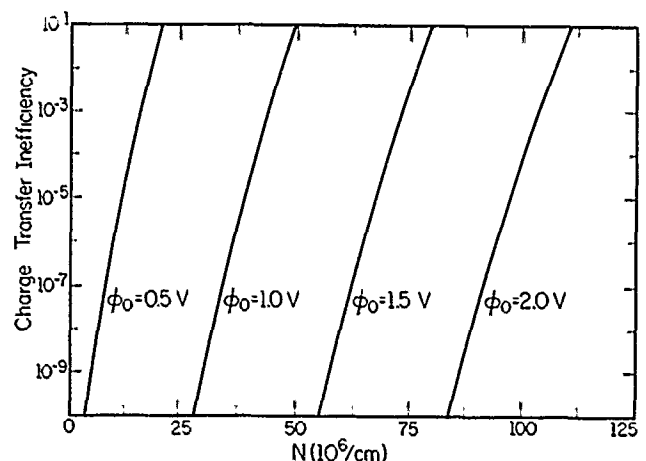


Fig 3. Variation of room temperature transfer inefficiency with charge capacity.

B. Nondestructive Charge Sensing

An attractive attribute of the ACT device is that the transport process within the channel is relatively immune to small perturbations in the surface electrical boundary condition. As a result, isolated surface electrodes can be used to sense the signal charge packets non-destructively via capacitive coupling. As a charge packet is transported under the Schottky channel plate it induces an "image" charge in the plate which moves in synchronism with the packet. Nondestructive charge sensing is accomplished by detecting the presence of the image charge as it moves along the channel plate without disrupting the transport of the charge packets.

Two basic types of nondestructive sensing (NDS) modes are depicted in Fig. 4. In the current sensing mode depicted in Fig. 4a, a small gap is introduced into the channel plate and the image current is detected by a resistive load connected across this gap. In the voltage sensing mode illustrated in Fig.

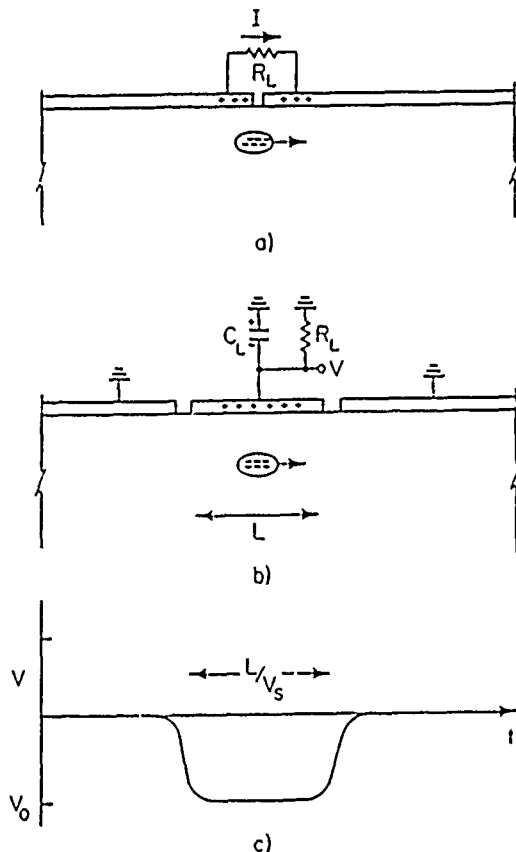


Fig 4. Non-destructive charge sensing.

- a) Current sensing
- b) Voltage sensing
- c) Impulse response for the floating electrode structure

4b, an isolated sense electrode with a capacitive load integrates the image current to produce an output voltage in direct proportion to the charge residing under the electrode. The impulse response of a voltage sensing electrode has a duration equal to the transit time of the packet under the electrode as shown in Fig. 4c. In practice this voltage is sensed by an amplifier with large input impedance R_i such that the $R_i C$ time constant is much longer than the transit time under the electrode.

An interesting aspect of the NDS operation involves the energy transfer mechanism. In particular, the signal energy dissipated in the load resistors of Fig. 4 is sourced by the SAW. For example, in the current sensing mode, the voltage induced across the resistor induces a small potential hill in the channel which the charge packet must climb. The work that the SAW expands to push the packet up this hill corresponds to the energy dissipated in the load resistor. The sense energy sourced by the SAW is a negligibly small fraction of the total wave energy and in principle, thousands of sense operations can be performed without significant SAW attenuation.

The charge packet flux line spreading (or fringing field) associated with the finite packet to channel plate separation creates an image charge distribution which has a spatial spread greater than the charge packet itself. This creates a very predictable rise time limitation to the impulse response of a NDS electrode which in turn, tends to reduce the detection bandwidth somewhat compared to the very fast destructive charge extraction process occurring at the output contact of the device. A more detailed description of the electrical circuit models, impulse and frequency response characteristics, and detection sensitivities of ACT NDS structures may be found in References 8 and 9.

The single tap NDS structures described here form the basis for multiple-tap ACT delay lines. These structures are repeatedly placed in the channel to form NDS electrode arrays which realize a specific transversal filter function. A variety of tap weighting schemes similar to those used in SAW and CCM filters may be implemented to achieve a specific array impulse response.

C. Channel definition and isolation.

The requirements for electrical isolation of an ACT transport channel are similar to those encountered in conventional GaAs CCDs or MEFETs. Early ACT devices³ were constructed using the mesa etch channel isolation method. Although this technique results in acceptable channel electrical isolation, the surface morphology variation at the mesa edge complicates high resolution photolithographic pattern definition in these regions. A simple and useful technique for device isolation which retains a planar surface morphology during processing is based on the use of a Schottky barrier guard rail which completely surrounds the desired active region. When this guard rail is reverse biased to achieve

deep depletion of the active layer underneath it, an electrostatic potential barrier is established. This barrier provides electrical isolation between the channel active layer and the epitaxial layer on the rest of the chip much like the isolation obtained between the source and the drain in a MESFET with a completely pinched off gate bias condition.

The guard rail architecture was initially demonstrated in the conventional GaAs CCD¹⁰ and subsequently applied to the simple ACT devices with good success⁶. However, the guard rail architecture is difficult to implement in more complicated ACT devices due to the difficulty associated with routing electrical connections to surface structures within the channel regions.

Recent ACT device architectures have been based on the use of proton isolation¹¹ for transport channel definition and isolation. A primary advantage to this approach is that the electronic isolation is relatively transparent acoustically; this tends to decouple the electronic and acoustic device design constraints. In addition, parasitic capacitances are minimized at the input and output nodes of the device, permitting larger gain-bandwidth products to be realized in the interface circuits which must provide an electrical impedance match between the high impedance ACT device and the relatively low impedance of a 50 ohm system. It has been found that the proton isolated ACT architecture permits a large degree of design flexibility while retaining excellent charge transfer performance.

III. ACT Device Performance

An overhead illustration of a typical proton isolated ACT delay line is shown in Fig. 5. A uni-directional SAW transducer, consisting of an interdigital transducer and a grating reflector generates large amplitude SAW with 8 μ m wavelength at an effective clock frequency of approximately 360 MHz. The ACT channel, shown as the shaded region in Fig. 5, is defined and isolated by a proton isolation implant which renders the epitaxial layer outside of the channel region semi-insulating. The delay line consists of an input section, a delay section and an output section all of which are illuminated by the SAW. Part of the transport channel plate is broken up into individual sections which form capacitive attenuators to reduce direct signal feed from input to output.

The response of an ACT delay line to a 50 MHz RF tone burst is shown in the oscillograms of Fig. 6. This delay line is 375 wavelengths or equivalent transfers long resulting in a delay time of 1.04 μ s. Fig. 6b shows the detail of the input and output waveforms using the delayed sweep function of the oscilloscope where the excellent waveform reproduction of the device is clearly evident.

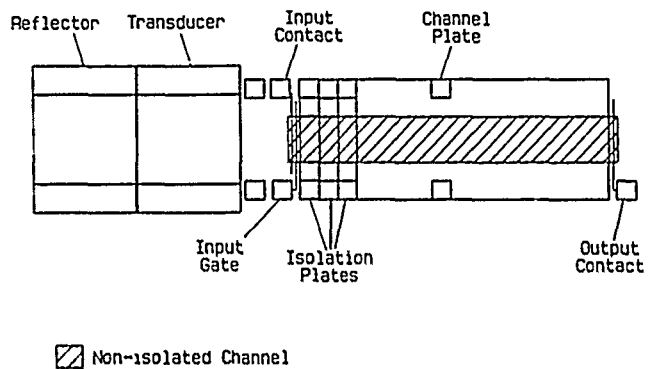


Fig 5. Overhead view of a proton isolated ACT delay line.

The frequency response of the ACT delay line is demonstrated in Fig. 7 using a tracking generator and spectrum analyzer test set. The device response is flat to within 0.4 dB over the entire Nyquist limited bandwidth of 180 MHz. The 0.4 dB rolloff at the Nyquist limit and the device transfer length are used to calculate¹² the transfer efficiency of the device at 0.99994. This number represents a worst case calculation since frequency response limitations at the input and output can contribute to the rolloff. To our knowledge, this is the highest transfer efficiency ever reported for any CCD operating with clock rates in the UHF band.

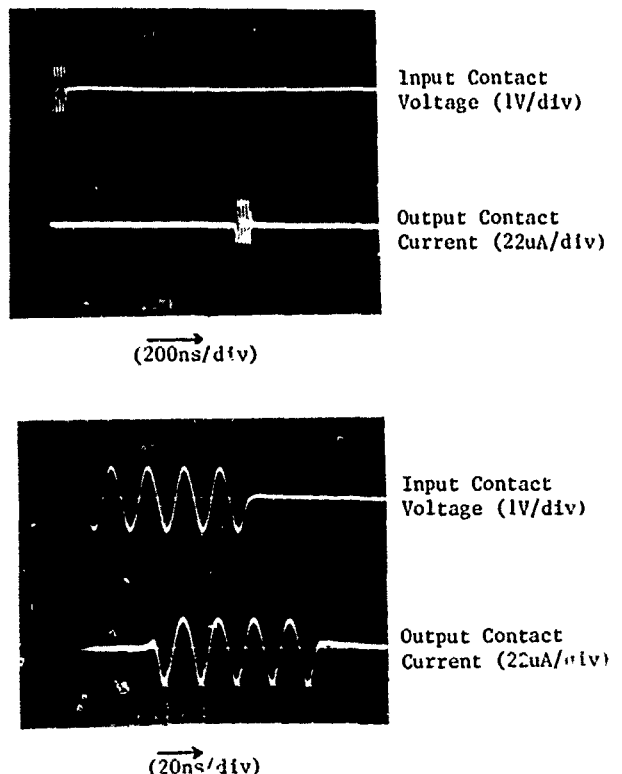


Fig 6. Proton isolated device response to 50MHz tone burst.

- a) Input and output waveforms
- b) Response detail

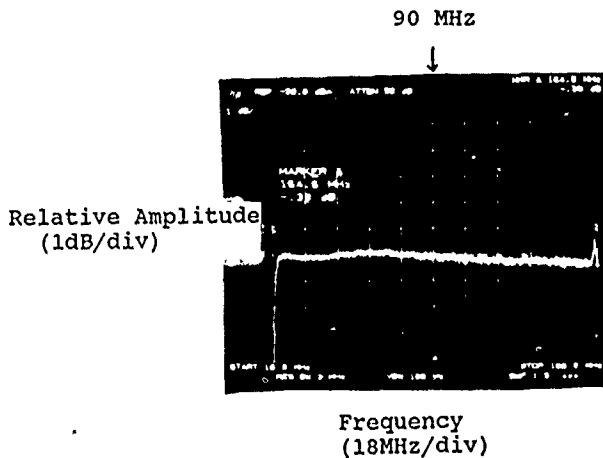


Fig 7. Frequency response of the proton isolated ACT delay line.

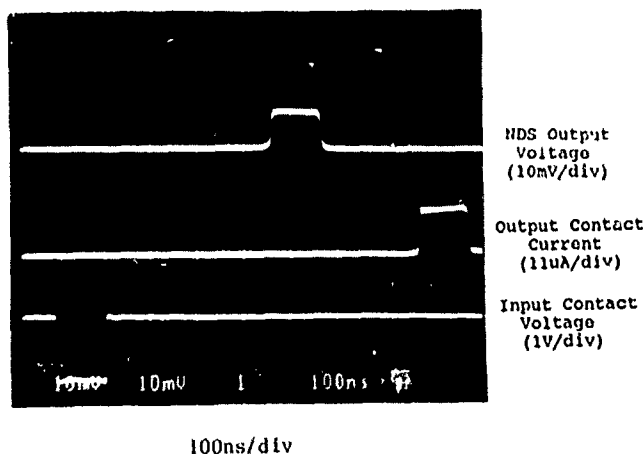


Fig 8. Destructive and nondestructive charge sensing pulse responses.

Figure 8 shows a comparison of single tap nondestructive and destructive charge sensing pulse responses obtained in a guard rail architecture ACT delay line operating at a SAW frequency of 360 MHz. This particular device has two outputs: a NDS electrode located 465 ns away from the input contact and a destructive sense output contact at the end of channel which has a total delay of 780 ns. The NDS electrode drives the gate of a GaAs FET which functions as an output buffer amplifier. It can be seen that the NDS output attenuates the small high frequency ringing component in the input signal due to its reduced rise time as compared to the ultra-fast destructive sensing output. The slight decay in the NDS output pulse amplitude is due to a low pass R-C time constant in the capacitively coupled output amplifier.

The frequency response of a simple ACT transversal filter based on the proton isolated architecture is shown in Fig. 9. This filter was designed as a diagnostic device in a family of 63 bit pseudo-noise (P-N) code correlators where all 63 bits have been given a weight of "one". In this case, the spread spectrum frequency response collapses down to a simple $(\sin f)/f$ type function with a fractional bandwidth of $1/63$. The NDS electrode array drives a single GaAs FET output buffer amplifier which primarily provides an impedance transformation from the low node capacitance of the array down to a 50 ohm system. No attempt was made to provide an impedance match to the high input impedance of the device ($\approx 30K\Omega$) which results in an input loss of approximately 28dB when the input VSWR is minimized with a shunt 50 ohm resistor. As can be seen from Fig. 9, the insertion loss of this device is 11 dB at a center frequency of approximately 82 MHz. If the device input had been matched, the filter would have produced a net gain, due in part to the large current gain of the NDS electrode array.

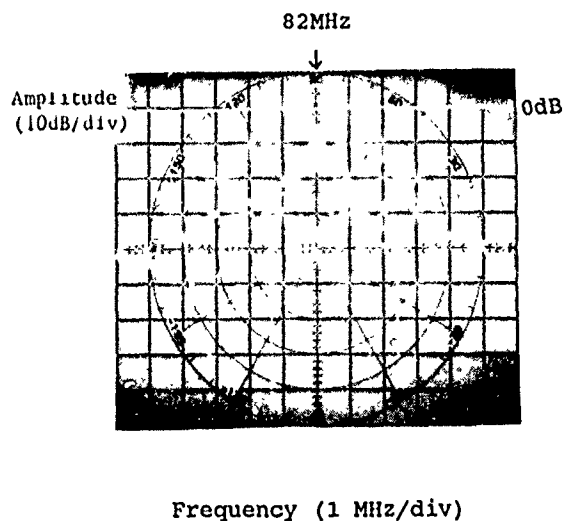


Fig 9. ACT transversal filter frequency response.

IV. Conclusion

The ACT device is a new type of high speed charge transfer device in which precise electron transport at the sound velocity is accomplished using the traveling wave electric field of a single frequency SAW. The device is constructed monolithically in GaAs and utilized to achieve tapped delay functions for signal processing applications. The basic operational principles and attributes of the ACT delay line are summarized and experimental results are given which demonstrate the high transfer efficiency, low interference charge detection and large signal bandwidths achievable in ACT delay lines based on a proton isolated architecture.

The nondestructive charge sensing capabilities which form the basis for multiple-tapped delay lines are discussed and the operation of an ACT transversal filter is demonstrated.

Acknowledgement

This work was supported jointly by the Defense Advanced Research Projects Agency and the Rome Air Development Center under contract numbers F30602-84-C-0126 and F30602-85-C-0170.

References

- 1) T.W. Grudkowski, G.K. Montress, M. Gil-
den and J.F. Black, "Integrated Circuit
Compatible Surface Acoustic Wave Devices
on GaAs," IEEE Trans. Microwave Theory
and Tech., MTT-29(12), p. 1348, Dec.
1981.
- 2) I. Deyhimi, R.C. Eden and J.S. Harris,
Jr., "GaAs and Related Heterojunction
Charge Coupled Devices," IEEE Trans.
Electron Devices, ED-27(6), p. 1183,
June 1980.
- 3) M.J. Hoskins, H. Morkoc, and B.J. Hun-
singer, "Charge Transport by Surface
Acoustic Waves in GaAs," Appl. Phys.
Lett., 41(4), Aug. 15, 1982.
- 4) M.J. Hoskins and B.J. Hunsinger,
"Monolithic GaAs Acoustic Charge
Transport Devices," 1982 IEEE
Ultrasonics Symposium Proceedings, cat.
No. 82CH1823-4, 1982.
- 5) M.J. Hoskins and B.J. Hunsinger, "Buried
Channel Acoustic Charge Transport
Devices in GaAs," 1983 GaAs I.C. Sym-
posium Technical Digest, IEEE Cat. No.
83CH1876-2, Oct. 25-27, 1983.
- 6) M.J. Hoskins, E.G. Bogus, and B.J. Hun-
singer, "Experimental Performance of the
Buried-Channel Acoustic Charge Transport
Device," IEEE Electron Device Lett.,
EDL-4(11), pp. 396-399, No. 1983.
- 7) M.J. Hoskins and B.J. Hunsinger, "Simple
Theory of Buried Channel Acoustic Charge
Transport in GaAs," J. Appl. Phys.,
55(2), pp. 413-426, Jan. 1984.
- 8) M.J. Hoskins, "Acoustic Charge Transport
in Gallium Arsenide," Ph.D. thesis,
University of Illinois, 1983.
- 9) C.W. Warren, "Nondestructive Sensing in
Buried Channel Traveling Wave Charge
Transfer Devices," M.S. thesis, Univer-
sity of Illinois, 1985.
- 10) M.D. Clark, C.L. Anderson, R.A. Jullens
and G.S. Kamath, "Planar Sealed-Channel
Gallium Arsenide Schottky-Barrier Charge
Coupled Devices," IEEE Trans. Electron
Devices, ED-27(6), p. 1183, June 1980.
- 11) D. D'Avanzo, "Proton Isolation for GaAs
Integrated Circuits," IEEE Trans.
Electron Dev., ED-29, No. 7, P. 1059,
July 1982.
- 12) Michael F. Tompsett and Edward J.
Zimany, Jr., "Use of Charge-Coupled
Devices for Delaying Analog Signals,"
IEEE J. Solid State Circuits, SC-8(2),
p. 151, April 1973.

FAST CHANNELIZER--A NEW FREQUENCY SORTING TECHNIQUE*

J. H. Elliott, R. B. Stokes and K. H. Yen

TRW Electronic Systems Group
One Space Park, Redondo Beach, California 90278

SUMMARY

Frequency sorting has always been a critical requirement for electronic warfare systems. Various techniques have been attempted with limited success. These included crystal video receivers, Bragg receivers, surface acoustic wave (SAW) microscan receivers and SAW channelized filter banks. For the last few years, TRW has been working on a unique and novel approach called a SAWFAST (Focused Acoustic Spectral Transform) Channelizer. This device offers high instantaneous bandwidth, high dynamic range and maintains the phase information while sorting. This channelizer has the key features of planar SAW filter technology--it is rugged, reproducible, compact, and low in cost. This device is purely passive and does not require a power supply, laser or additional optics like Bragg cell.

This paper will discuss the design and performance of the SAWFAST Channelizer. First the basic principle of operation will be discussed. This operating principle is similar to that of an optical diffraction grating. Test results will then be presented for devices covering an input bandwidth of 160 MHz and 250 MHz. Typically, the insertion loss is 25 dB and sidelobe rejection >40 dB.

PRINCIPLE OF OPERATION

The SAWFAST Channelizer separates a wide input bandwidth into multiple narrowband output channels. Typical input fractional bandwidths are 50-60%, while output bandwidths vary from 0.5% to 2.5%. The filter is fabricated on a single crystal lithium niobate substrate. It consists of one input transducer and multiple output transducers as depicted in Figure 1. The input transducer is a surface acoustic wave implementation of an optical diffraction grating. This device was first described by Brooks and Wilcox [1]. Another device operating on a related principle was described by Lewis [2] at the 1985 Ultrasonics Symposium.

An optical diffraction grating consists of a large number of parallel line scatterers. White light is dispersed into its component colors for all diffracted orders except the zeroth order. This is a result of the wave nature of the light and phase interference phenomena. The diffraction order is simply in light paths between adjacent scatterers measured in wavelengths. The SAWFAST input transducer consists of a number of small sub-transducers equally spaced in an approximately lateral array along the substrate. The array of sub-transducers is also curved, focusing the waves to a point whose position indicates the wavelength. The input transducer is also the source of the surface acoustic waves. It therefore functions as the source, grating and lens of an analogous optical spectrograph.

The SAWFAST Channelizer uses the first diffracted order. For any particular wavelength, the distance from the focus spot to successive input

sub-transducers differs by one wavelength. Since the SAW wavelength is inversely proportional to frequency, the device is a frequency channelizer. The acoustic signals present at the focal plane are converted back into electrical form by an array of output interdigital transducers. These transducers are centered at each output channel center frequency and placed at the corresponding focus spot. Additional channel shaping is provided by output transducer weighting.

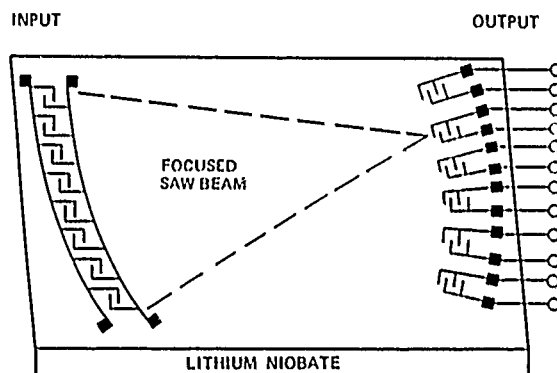


Figure 1: SAWFAST Channelizer physical layout.

SAWFAST Channelizers have been built with instantaneous bandwidths up to 500 MHz. The input fractional bandwidth is limited to $\leq 60\%$, beyond which the diffracted orders overlap. The number of output channels is limited by size constraints at the focal plane and output channel crosstalk considerations. Devices have been built with 8 to 27 output channels. The total number of channels can be increased by using several devices with interleaving channels. The device performance presented in the next section describes two such channelizers.

DEVICE PERFORMANCE

A photograph of a prototype SAWFAST Channelizer is shown in Figure 2. The SAWFAST Channelizer shares features with all SAW filters - small size, rugged and passive. This device channelizes a 160 MHz input band into 16 output channels. The output channels have 3 dB bandwidths of 5 MHz and a channel-to-channel separation of 10 MHz. This device is one of a pair of devices designed to cover the 160 MHz band with 32 channels.

The measured frequency response of all sixteen channels is plotted in Figure 3. The minimum insertion loss with the input transducer tuned is 25 dB. Since the filter is passive, its single-tone dynamic range is limited by the noise floor at ~ -90 dBm. The input power is limited to $\sim +20$ dBm to prevent

*This work was partially supported by the Naval Research Laboratory under contract N00014-81-C-2483.

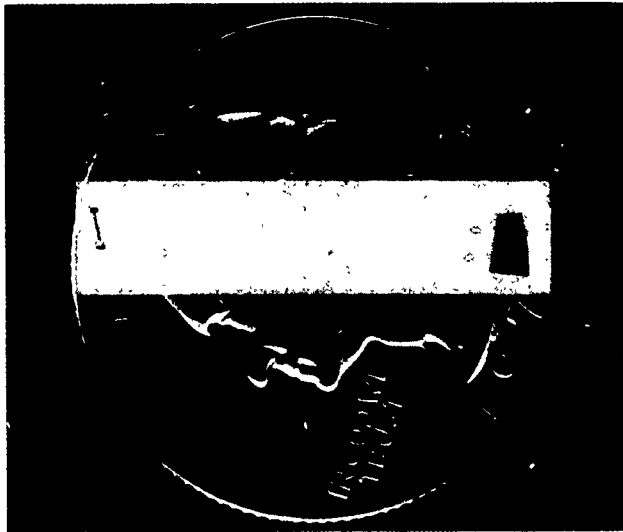


Figure 2: SAWFAST prototype device which separates an input 160 MHz bandwidth into sixteen output channels.

transducer damage. This implies a single tone dynamic range of 85 dB. The device is highly linear so that the two-tone intermodulation products are typically suppressed by more than 60 dB. The frequency response of two center channels is plotted in Figure 4. The sidelobe rejection of the best channel is 37 dB. The two-tone dynamic range is therefore 37 dB. The output transducers are weighted using a withdrawal technique to improve the channel shape.

Since the FAST Channelizer is limited in fractional bandwidth, the center frequency must be increased to increase the instantaneous bandwidth. The upper limit on bandwidth is set by the lithographic resolution limits. With modern photolithography, this implies a FAST bandwidth maximum of ≈ 600 MHz for a conventional transducer design. The scalability of FAST design concepts has been shown by printing a device at 1X, 5X and 10X reduction factors corresponding to 50 MHz, 250 MHz and 500 MHz bandwidth. The peak insertion loss changed by less than 2 dB. The passband shape was essentially unchanged by the scaling. As a result, recent efforts have concentrated on 250 MHz bandwidth designs centered at 500 MHz.

The frequency response of one such device is plotted in Figure 5. Again, this device is an every other channel design with 10 MHz channel bandwidth and 20 MHz channel separation. The output transducers provide good channel shaping through withdrawal weighting. The frequency response of one channel is shown in Figure 6. The close-in sidelobe levels are 40 dB down, and the far away sidelobes ≈ 48 dB down from the peak signal. These measurements were performed using a 50 ohm network analyzer without matching circuits. Tuning the input would decrease the loss by 3 dB, to 25 dB. Tuning the output channels is not necessary.

CONCLUSIONS

The FAST Channelizer is a surface acoustic wave filter that separates a wide input band (50-600 MHz) into multiple output channels. Its operating principles are analogous to those of an optical diffraction grating. The two devices described here cover 160 MHz and 250 MHz of bandwidth. Each is one of a pair of devices designed to cover the input bandwidth with no gaps. Typical tuned insertion losses are 25 dB. In the best channels, sidelobe rejection is ≈ 40 dB, promising excellent dynamic range in the presence of strong signals.

Future EW receivers will incorporate some type of channelizer for threat identification. The SAWFAST Channelizer is a simple, rugged, small device which offers adequate bandwidth and resolution for many system applications.

REFERENCES

- [1] R. E. Brooks and J. Z. Wilcox, "SAW RF Spectrum Analyzer/Channelizer Using a Focusing Phased Array Transducer," 1985 IEEE Ultrasonics Symposium Proceedings, Paper L-1
- [2] M. F. Lewis, "A Compact and Efficient SAW Filterbank," 1985 IEEE Ultrasonics Symposium Proceedings, Paper X-1

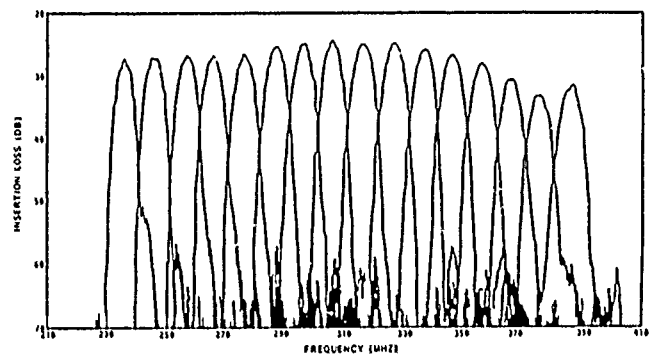


Figure 3: Frequency response of all sixteen channels of the 160 MHz bandwidth prototype SAWFAST Channelizer.

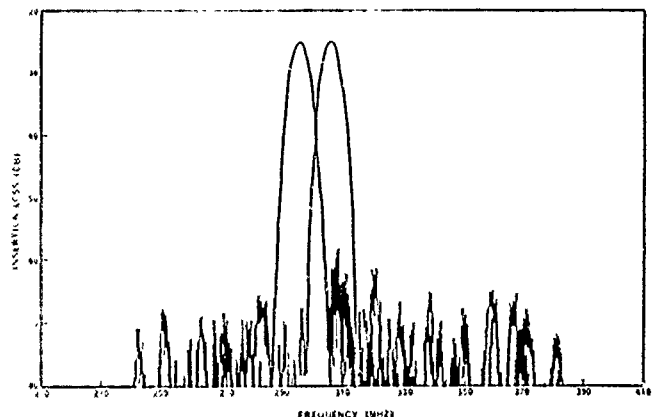


Figure 4: Frequency response of the two center channels of the 160 MHz bandwidth prototype SAWFAST Channelizer.

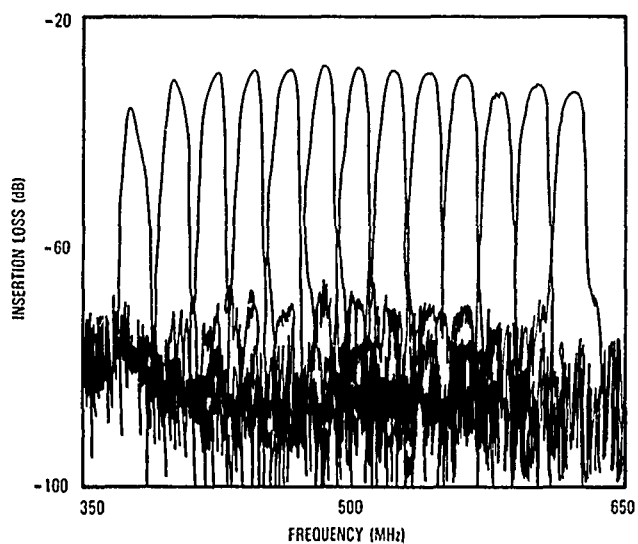


Figure 5: Frequency response of a 250 MHz bandwidth SAWFAST Channelizer.

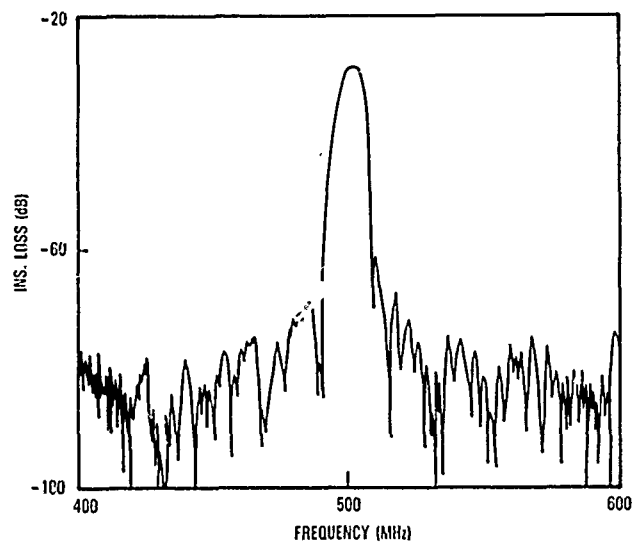


Figure 6: Single channel frequency response of the 250 MHz bandwidth SAWFAST Channelizer greater than 43 dB sidelobe rejection.

GROUP DELAY MEASUREMENTS - A SENSITIVE METHOD FOR
DETECTING SPURIOUS CRYSTAL RESONANCES

FERDINAND K. EULER, Rome Air Development Center
Solid State Sciences Division
Hanscom AFB, MA 01731-5000

Abstract

Group delay, or the derivative ϕ' of phase with respect to frequency, is being applied to the problems of detecting weak spurious crystal modes and determining the origin of observed peaks in the resonance resistance vs temperature curve.

Crystal resonators were placed in an oven and connected to a programmable network analyzer and synthesizer. Automated measurements were taken with the temperature ramped at rates of 0.1 and 0.3°C/min. 1500 phase measurements were made while the frequency f was swept through a 15 kHz range. The sweep was repeated 40 times for each selected temperature, the phase values averaged, differentiated and $\phi'(f)$ plotted.

Arrays of $\phi'(f)$ curves were plotted at 5°C temperature intervals, and show nanosecond peaks of spurious modes moving through the chosen frequency range as the temperature increases. Their peak heights grow as they approach the main resonance. The resonance resistance shows peaks at the temperatures of intersection.

Data taken on SC- and AT-cut quartz resonators will be discussed. We also describe measurement strategies such as the choice of Δf to generate $\phi' = \Delta\phi/\Delta f$, the number of repetitive scans needed to lift the signal out of the noise and the use of running averages of ϕ before forming the derivative.

Introduction

One way to characterize precision quartz resonators is the measurement of the resonance parameters as function of temperature T .¹⁻³ In particular, the resonance resistance R varies with T and may show one or more peaks of widely varying strength. These variations are undesirable as they may lead to instabilities of amplitude (activity dips) or frequency when the resonators are deployed in oscillators.⁴ The $R(T)$ peaks are generally attributed to coupling of the crystal's main resonance with spurious modes.^{4,5} Detection of these modes and measurement of their frequency-temperature characteristics has been difficult and time consuming.⁶ But with the availability of programmable network analyzers, these measurements can be automated. Routinely, $R(T)$ is determined from automatically repeated frequency sweeps through the series resonance of the crystal as the temperature is slowly ramped through the range of interest.³ We extended this method to the measurement of mode spectra as function of temperature,⁷ plotting signal magnitude vs frequency f for a series of selected temperatures. The sensitivity of this technique is limited by the amplitude resolution of the network analyzer and by masking of spurious modes by the main resonance. We have now found that these limitations can be overcome by measuring the phase ϕ rather than the magnitude A of the output signal and plotting the derivative $\phi' = d\phi/df$ vs f . The negative phase slope $-\phi'$ is widely known

as group delay which is an important filter parameter⁸ and directly measurable with advanced network analyzers.⁹ Recently group delay measurements have been utilized for analyzing the resonance characteristics of precision quartz resonators.¹⁰

This paper presents the application of ϕ' scanning to the problem of detecting weak spurious modes. After the experimental details, $A(f,T)$ scanning and $R(T)$ data are discussed before introducing ϕ' scanning and its advantages. Further improvements gained by averaging are shown next, and the paper concludes with the application to two examples of precision AT-cut resonators.

Experimental

Our quartz resonators are precision 5MHz, 5th overtone resonators, AT or SC cut, and were placed in an oven. Temperature control capabilities include ramping with adjustable rates. The resonators were connected with coaxial wiring through a π network to a programmable network analyzer, Hewlett-Packard (HP) model 3570, operated with 10 Hz measurement bandwidth, digitized with 0.01 dB amplitude and 0.01° phase resolution. A programmable synthesizer, HP model 3330, generated r.f. voltage with -10 dBm output ampli-

Table I. Experimental Parameters

measurement type	A or $\phi'(f,T)$	R(T)
dT/dt (°C/min)	0.1	0.3
T interval (°C)	5	~0.3
π -network termination (Ω)	50	4.7
series resistor (Ω)	-	47
frequency sweep step (Hz)	40, 10	>0.1*
number of steps	1500	100
time/step (ms), weak modes	1	1
strong modes	-	100

* automatically adjusted for resonance linewidth

tude and 0.1 Hz minimum sweeping step. Minimum stepping time was nominally 1 ms but slowed down to effectively ~20 ms by the software. The measurement process was controlled by a desk top computer via IEEE-488 interface bus, and the data stored on magnetic tape. Table I shows the different parameter selections for $A(f,t)$, $\phi'(f,T)$ and $R(T)$ measurements.

Magnitude Scanning

Figure 1 shows a series of plotted magnitude scans $A(f)$, each consisting of 1500 measurements of the magnitude A of the signal received by the network analyzer, spaced 40 Hz apart. The scans were repeated every 5°C and show the frequency-temperature characteristics of four crystal resonances, labeled A300, A301, A302 and X. The strongest mode of these is A302, and its magni-

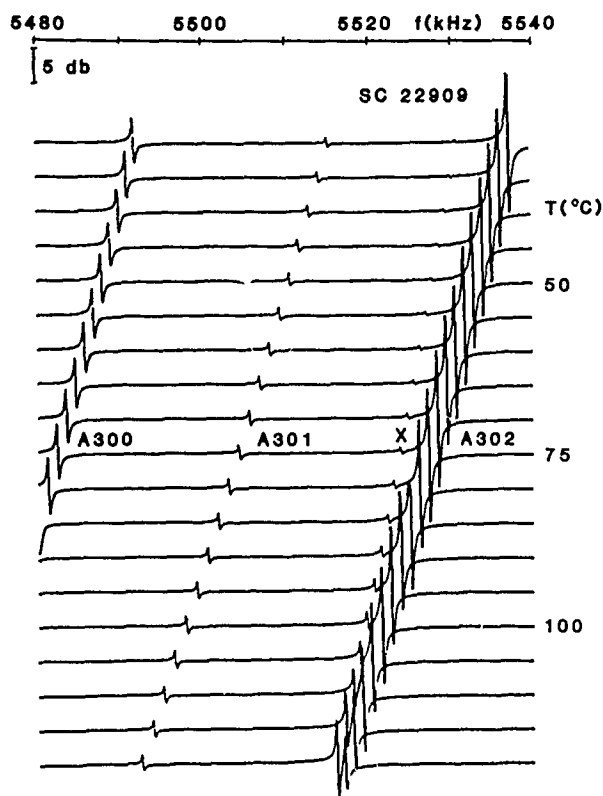


Fig. 1. Example of magnitude scanning $A(f, T)$, showing three A modes of a 5 MHz, 5th overtone SC-cut quartz resonator (SC 22909).

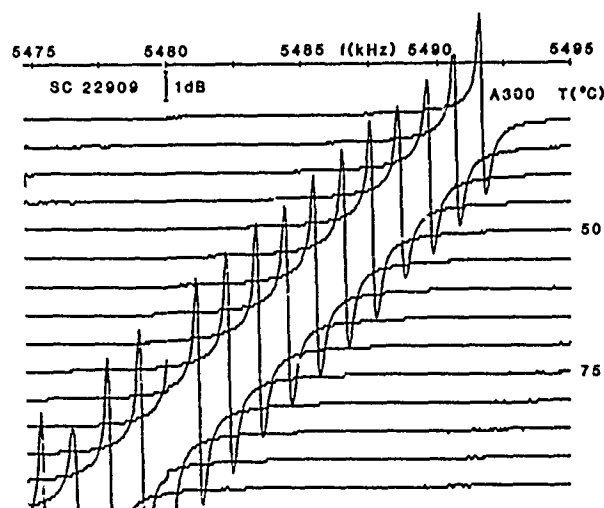


Fig. 2. Magnitude scanning $A(f, T)$ of crystal SC 22909, detailing mode A300 on expanded scales.

de variation is due to coupling with X. Mode A300 seems rather normal in the $A(f)$ scans. The background between the modes appears flat, except for occasional jogs of 0.1 dB caused by rounding.

Magnitude scans were also made with 10 Hz frequency stepping for details of mode A300. The result is shown in Fig. 2. We note that the background changes in steps of 0.1 dB, due to measures to cope with the limitations of our magnetic tape capacity. In order to record all

scans of one run on one track of the cartridge, the measured A values were compressed into three digits, omitting the 0.01 dB digit. The rows of jogs paralleling the A300 mode are part of the ends of its tails and not independent weak modes. We note further that opposite tails are 0.2 dB or further apart over a stretch of 8 kHz. That makes it difficult to detect weak modes close to a major resonance by magnitude scanning. All we can see of the A300 mode is a number of dips in the peak height located at 40, 45, 65 and 100 °C (the interruption in the trace of the 85 °C scan comes from splicing data of two runs, at 5480 kHz).

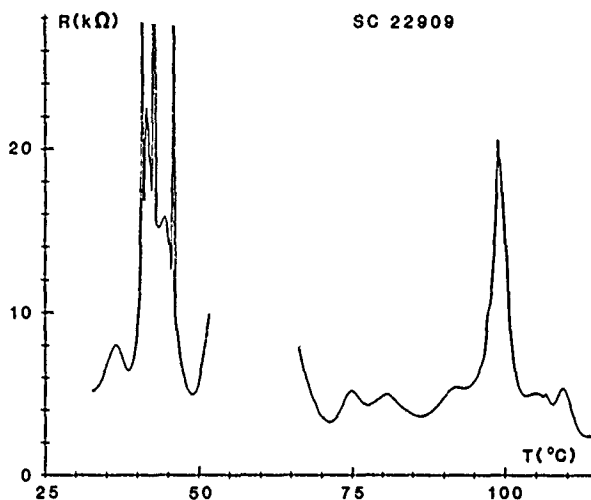


Fig. 3. $R(T)$ plot of mode A300 of resonator SC 22909. Note loss of resonance between 50 and 65 °C.

Mode Spectra and $R(T)$

Figure 3 shows measured $R(T)$ of the A300 mode. One notices two regions of major peaks located at 40-45 and 95-100 °C. These correspond clearly to the dips of the $A(f)$ peaks (in Fig. 2) at 40, 45 and 100 °C. The loss of resonance between 50 and 65 °C was caused by high resistance also, corresponding to the $A(f)$ peak dip near 65 °C. We expect this as R was calculated from the maximum measured signal magnitude. The generally accepted^{4,5} cause of the $R(T)$ peaks is coupling of the mode of interest (here A300) with weak spurious modes. In this and other cases the spurious modes are too weak to be detected by magnitude scanning.

Phase Slope or Group Delay Scanning

In Fig. 4 phase slope $\phi'(f)$ is plotted as calculated from the measured phase ϕ according to

$$(1) \quad \phi' = (\Delta\phi / 360^\circ) / \Delta f$$

with $\Delta\phi = \phi_2 - \phi_1$ and $\Delta f = f_2 - f_1$ chosen to be 200 Hz (the interval between 20 consecutive measurements). The scans show a number of very weak modes. Those near A300 are filled in solid and connected with trend lines. The mode A300 is superimposed in a different scale and shaded. The strongest coupling of two crossing modes occurs near 100 °C and appears to be the reason for the sharp $R(T)$ peak at 98 °C. Other mode

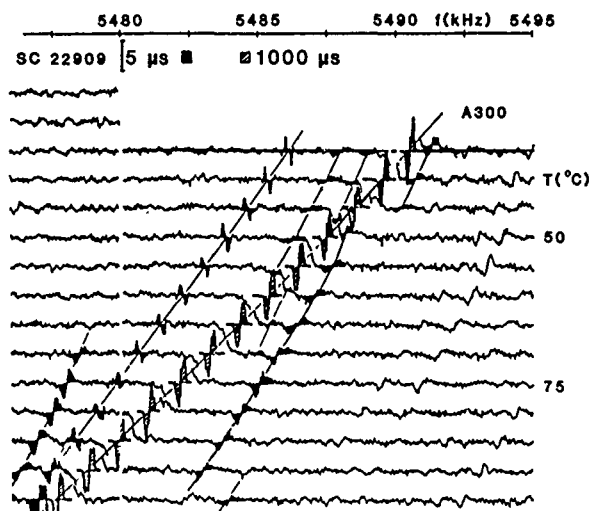


Fig. 4. Phase slope $\phi'(f,T)$ scanning of resonator SC 22909, covering mode A300.

crossings occur between 45 and 60°C but their relationship with the $R(T)$ peaks in that region is not so clear.

Nevertheless, one can conclude that $\phi'(f)$ plots are quite practical for detecting weak resonance modes not detectable by $A(f)$ scans. There exists correlation between these modes and the position of $R(T)$ peaks.

Negative ϕ' is known as group delay, and the sign reversal makes the group delay signatures of crystal resonances appear like magnitude resonance curves. Like extrema in magnitude and group delay correspond to each other as the crystal admittance traces a circle in the complex number plane when the frequency passes through resonance (the diagrams and discussions of this paper are in terms of ϕ' rather than group delay).

Table II illustrates the sensitivity of $\phi'(f)$ and $A(f)$ scanning on two examples, peaks A301 and X of Fig. 1. The analyzer's phase resolution of 0.01° leads to the ϕ' resolutions shown as function of the selected

Table II. Sensitivity Comparison

	resol- ution	A301 peak	s/r	X peak	s/r
	(μ s)	(μ s)		(μ s)	
ϕ' with $\Delta f = 10$ Hz	2.8				
50 Hz	0.56				
200 Hz	0.14	280	2000	130	930
	(dB)	(dB)		(dB)	
A as measured	0.01	0.8	80	0.2	20
plotted	0.1		8		2

s/r signal-to-resolution ratio

differentiation interval Δf . Note that one can improve the resolution by increasing Δf , limited by the amount of distortion

one can tolerate. Generally, increased Δf causes an apparent widening of the frequency interval between maximum and minimum ϕ' . This, however, is unimportant if one is merely interested in detecting a weak mode. So, with $\Delta f = 200$ Hz, one can detect A301 and X with excellent signal-to-noise ratio, even without repeating and averaging of $\phi'(f)$ scans. In contrast, $A(f)$ scans are limited by the analyzer resolution of 0.01 dB without adjustment possibilities.

Measured by the signal-to-resolution ratio s/r, ϕ' -scanning yields a sensitivity gain of 25 to 40 over A-scanning. A factor 7 of these gain values comes from the superior phase resolution of the analyzer, as the magnitude resolution of 0.01 dB corresponds to a phase change of $\approx 0.07^\circ$. The remaining portion of the sensitivity gain (factor 3 to 6) must come from the differentiation.

Differentiation is used in many different measurement techniques, e.g. electron spin resonance, for enhancing small differences and suppressing unneeded base values. As we have seen in the case of magnitude scanning, carrying the base values led to the decision to sacrifice available measurement resolution for data storage economy. With ϕ' scanning, the unneeded phase background is suppressed to zero, and three-digit data storage can handle 10 orders of magnitude in $\pm \phi'$ peak values with 4.7% relative precision, without loss of resolution.

One final advantage of ϕ' -scanning is the narrowness of the resonance modes. Thus one is able to detect spurious weak resonances as close as 1 kHz next to a major mode, as shown in the lower left corner of Fig. 4.

Refinement by Averaging Over Repeated Scans

The plot in Fig. 4 was generated by single scans for each selected temperature. With $\Delta f = 200$ Hz, considered here an apparent practical upper limit, and the chosen input amplitude together with the π -network resistances and the resonator's parallel capacitance (≈ 1 pF), the observed single-scan noise of ϕ' is ≈ 0.5 μ s. Single-scan mode detection limit is a few μ s peak to peak. Averaging over n repeated scans reduces these limits by \sqrt{n} , at the price of increased scanning time. Averaging, e.g., 40 scans yields ≈ 500 ns peak-to-peak detectability but requires $40 \times 1500 \times 20$ ms = 20 minutes. With 0.1°C/min ramping, the temperature changes over 2°C during the 40 scans. And 100 to 300 Hz/ $^\circ\text{C}$ mode shifting then causes line broadening of 200 to 600 Hz.

With these rather substantial scanning times, it pays to refine the software for minimum execution time, e.g. repeat and average the ϕ' -scans and calculate ϕ' from the average. One further step towards noise reduction is to compute a running average over a few percent of the length of the already averaged ϕ' scan before calculating ϕ' . This does not increase the scanning time but, if taken too far, will reduce the mode signals. This running average, therefore, was extended only over the interval Δf .

Figure 5 shows a set of ϕ' -scans, derived from averaged ϕ' measurements, of the region near the main resonance of SC 22909. On the increased

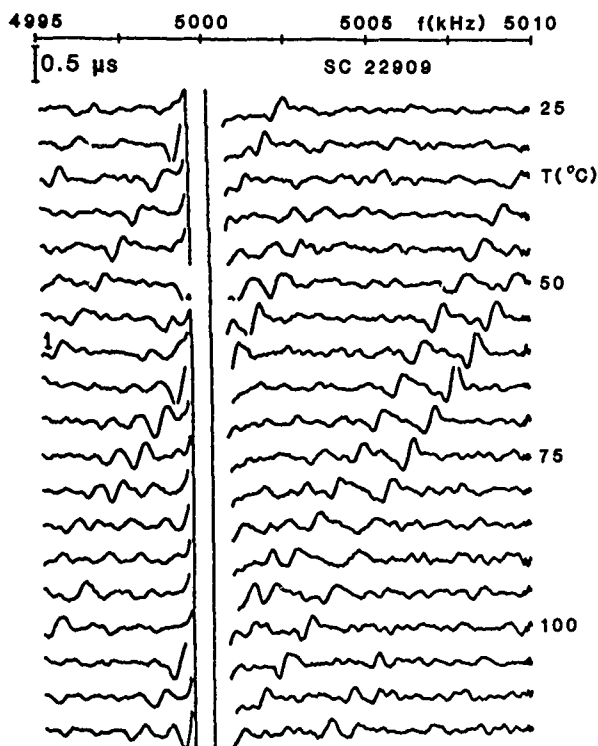


Fig. 5. Main resonance region, 5 MHz, 5th overtone C-mode of crystal SC 22909, $\phi'(f, T)$ scanning with $\Delta f = 200$ Hz, average of 40 phase scans for each temperature.

scale one sees ϕ' peaks of ≈ 100 ns height moving across the frequency range with T . When compared with Fig. 4, the background noise is greatly reduced, by averaging 40 scans and calculating running averages. The influence of the latter is missing at the upper f limit where one may see a slightly noisier trace over the last 200 Hz. Excursions of $|\phi'| > 0.25$ μ s have been eliminated from the plot for clarity. The main resonance is ≈ 4 ms peak-to-peak.

Figure 6 shows the corresponding $R(T)$ plot, and one can correlate, at least qualitatively, the T positions of the $R(T)$ peaks with mode crossing. A good example is the mode labeled 1 in Figure 5, crossing the main resonance between 25 and 30°C, just where the first $R(T)$ peak is found. Similarly, the $R(T)$ peak at 112°C has a corresponding mode crossing. The other $R(T)$ peak positions are hard to correlate to mode crossings. The strong $R(T)$ peak near 50°C cannot be correlated to any single mode crossing, but there are several mode crossings between 50 and 70°C. Further work is needed to clarify these details. For now, one can say in general that spurious modes with ϕ' peaks of 1 to 3 μ s may cause $R(T)$ peaks of 10 - 30 k Ω and ϕ' peaks of 100 - 300 ns lead to 100 - 300 Ω peaks.

Further Examples: AT-Cut

Figures 7 - 9 show examples of $R(T)$ and $\phi'(f, T)$ plots of two 5 MHz, 5th overtone AT-cut resonators from different batches but fabricated with the same design parameters.

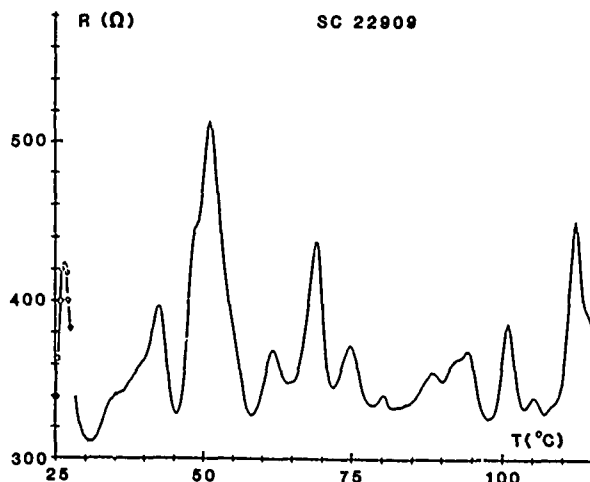


Fig. 6. Main resonance resistance $R(T)$ of crystal SC 22909.

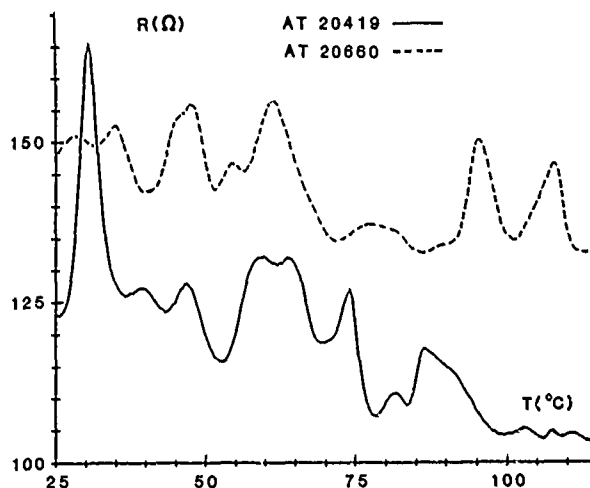


Fig. 7. $R(T)$ plots of resonators AT 20419 and AT 20660; the latter has been displaced for clarity: subtract 30 Ω to obtain the measured R values.

Figure 7 shows the $R(T)$ plots of the two resonators. Typical for AT-cut resonators, baseline R is near 100 Ω and the peaks extend 10 to 40 Ω above baseline. The differences between the two patterns are individual crystal differences that appear to be beyond the control of state-of-the-art crystal design ("process-related").

The corresponding $\phi'(f, T)$ scans are shown in Figs. 8 and 9, and these patterns too show the individual crystal differences. Crystal AT 20419 shows one single-mode crossing, corresponding to an isolated $R(T)$ peak at low temperature, and multiple crossings between 60 and 75°C related to bunched $R(T)$ peaks in the same region. In contrast, the mode crossings of AT 20660 (Fig. 9) are not well defined, and the $R(T)$ peaks are all rather weak.

There is one common feature in the two $\phi'(f, T)$ patterns: a relatively strong spurious mode near the lower right corner of Figs. 8 and 9. Its

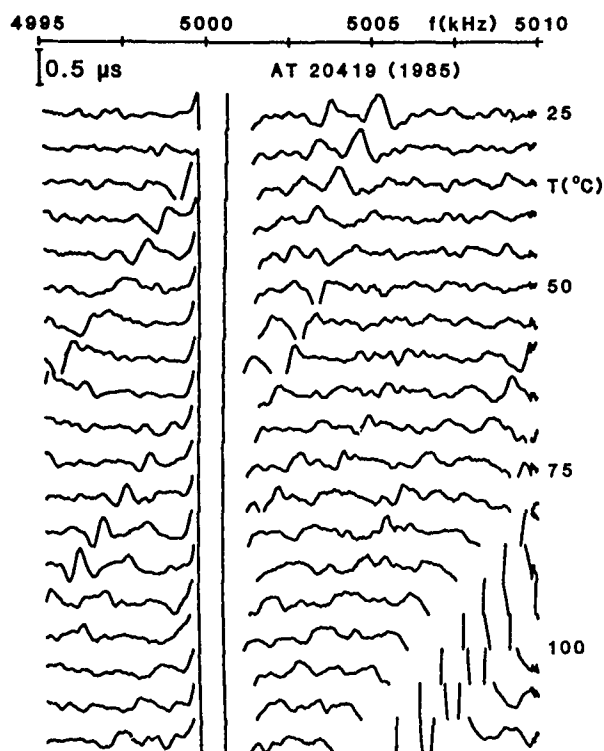


Fig. 8. Averaged $\phi'(f,T)$ scans of resonator AT 20419; $\Delta f = 200$ Hz; 40 scans averaged for each temperature, then running average over 200 Hz.

peak-to-peak strength is 10 and 50 μ s in AT 20419 and 20660, respectively. Its crossing with the main mode occurs near 140°C and leads to a strong R(T) peak there ≈ 100 Ω above baseline. This mode occurs in all crystals of the two batches and belongs to a series of 3rd overtone A-modes ("design-related"⁴). The ϕ' strength of the main mode is ≈ 5 ms peak-to-peak, for both crystals.

Conclusion

It has been shown that group delay or $\phi'(f)$ plots can be used to reveal the frequency-temperature characteristics of extremely weak crystal resonances. Formation of the derivative $\phi' = \Delta\phi/\Delta f/360^\circ$ eliminates the background values of ϕ and enhances its changes. With the choice of Δf one can control the measurement resolution. With proper noise reduction techniques, it is possible to detect modes with ϕ' peaks in the order of 100 ns. Their crossings with a major resonance correlate with the positions of R(T) peaks.

Acknowledgment

It is my wish to thank Mr. Alfred Kahan, Dr. A. Strauss and Dr. Arthur Ballato for encouragement, stimulating discussions and sharing of results before publication.

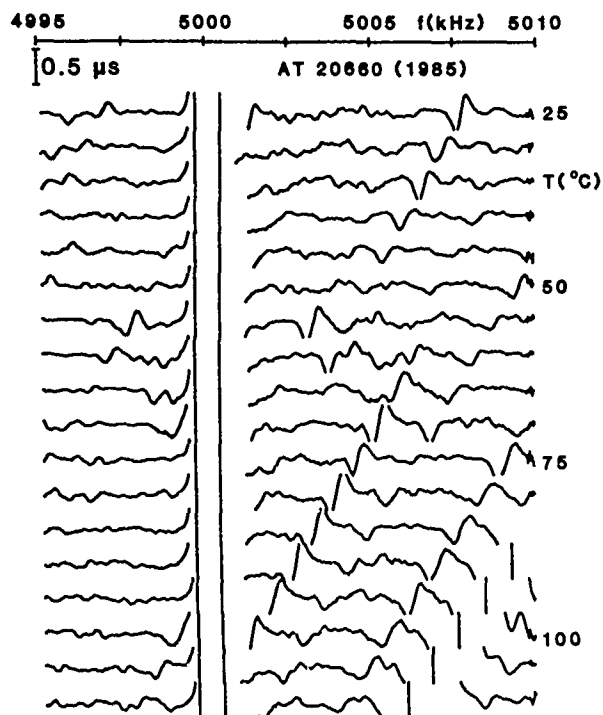


Fig. 9. Averaged $\phi'(f,T)$ scans of resonator AT 20660.

References

1. F. Euler et al., Proc. 32nd Annual Frequency Control Symposium (AFCS), 24 (1978).
2. H. G. Lipson et al., Proc. 35th AFCS, 329 (1981).
3. F. Euler et al., Proc. 36th AFCS, 115 (1982).
4. A. Ballato and R. Tilton, IEEE Trans. Inst. and Meas., IM-27, 59 (1978).
5. A. F. B. Wood and A. Seed, Proc. 21st AFCS, 420 (1967).
6. H. Fukuyo, H. Yoshie and M. Nakasawa, Bull. Tokyo Inst. Technology, 82, 53 (1967); also Proc. 21st AFCS, 402 (1967).
7. F. Euler and A. Kahan, Proc. 38th AFCS, 150 (1984).
8. F. Kuo, *Network Analysis and Synthesis*, 2nd Edn., J. Wiley & Sons, New York, 1966, Ch.8 Sec.5, 245.
9. For example, Hewlett-Packard models 3570 and 3577.
10. A. Strauss, private communication.

RF SPECTRUM OF THE OSCILLATOR SIGNAL UNDER NON-STATIONARY PHASE INSTABILITIES

by

Berni JOSS, Laurent-Guy BERNIER and Fred GARDIOL
Laboratoire d'Electromagnétisme et d'Acoustique
Ecole Polytechnique Fédérale, Lausanne, Switzerland.

Summary

The RF Spectrum of the oscillator under non-stationary phase instabilities is discussed. The power spectrum is shown to be dependent upon the frequency fluctuations. Assuming that the one-sided power spectral density of the reduced frequency fluctuations is a single term of the polynomial model and has the form $S_{yy}^+(f) \equiv h_\alpha f^\alpha$, three cases are shown to exist. If $\alpha \geq 0$ the power spectrum is stationary, if $-1 \leq \alpha < 0$ it is quasi-stationary and if $\alpha < -1$ it is non-stationary. Non-stationary frequency fluctuations are modelled by fractional integration of a white noise process. Besides, a computer calculation of the power spectrum using an equivalent stationary phase noise process is presented that takes into account the duration of observation. Both methods are compared against experimental measurements.

Introduction

High capacity communication systems, high performance ranging and navigation systems, as well as coherent radars and VLBI experiments are all examples of frequency control applications that require the synthesis of a highly stable reference signal at microwave frequencies. Usually, both the frequency stability and the spectral purity of the reference determine the performances of such systems. The frequency stability is related to the random phase and frequency fluctuations of the carrier while the spectral purity is related to the RF power spectrum of the signal itself. It is common practice to characterize both the frequency stability and the spectral purity by a single measurement of $L(f)$ which is defined as the RF power spectrum translated about the origin and normalized by the carrier mean power. Assuming that the phase noise is a stationary random process, that its quadratic mean value is much smaller than unity and that the amplitude noise is negligible, it can be shown that $L(f)$ is equal to the two-sided power spectral density of the phase noise^{1,2}. However, as Baghdady² pointed out as early as 1965 the phase fluctuations in oscillators

are never stationary because flicker type noise sources always dominate at low Fourier frequencies. Hence the quadratic mean value of phase fluctuations as well as the power spectrum itself are undefined in a strict sense. Besides, the mean quadratic value of the phase noise over the duration of observation may exceed unity in a microwave oscillator or after frequency multiplication of a quartz reference oscillator. When this is the case, the spectrum starts to spread and there is no more a simple relationship between frequency stability and spectral purity. In this paper we study the relationship between spectral purity and frequency stability under general conditions, i.e. without assuming that phase fluctuations are necessarily very small and stationary.

Analytic Representation of the Oscillator

It can be shown that the phasor representation of the oscillator possesses all the properties of an analytic signal¹. Hence, in the following analysis, we shall make use without proof of several properties of analytic signals. Let the output signal $s(t)$ from the oscillator be the real part of the analytic signal

$$\psi(t) \equiv A \gamma(t) \exp(j2\pi\nu_0 t). \quad (1)$$

A is the nominal amplitude and the phasor $\exp(j2\pi\nu_0 t)$ represents the carrier at the nominal frequency ν_0 . The complex envelope of the analytic signal is given by

$$\gamma(t) \equiv (1 + \epsilon(t)) \exp(j\phi(t)) \quad (2)$$

where $\epsilon(t)$ is the amplitude noise and $\phi(t)$ the phase noise of the oscillator. Assuming that amplitude noise is negligible with respect to phase noise, the autocorrelation function of the complex envelope is

$$\begin{aligned} R_{\gamma\gamma}(t_1, t_2) &\equiv E\{\gamma(t_2)\gamma^*(t_1)\} \\ &= E\{\exp[j(\phi(t_2) - \phi(t_1))]\} \end{aligned} \quad (3)$$

where $E\{\}$ means the statistical expected value and $*$ the complex conjugate. It can be shown

that the one-sided power spectral density of the output signal from the oscillator is given by

$$S_{ss}^+(t_1, f) = A^2/2 S_{yy}(t_1, f - \nu_0) \quad (4)$$

where $S_{yy}(t_1, f)$ is the two-sided power spectral density of the complex envelope, i.e. the Fourier transform of its autocorrelation function :

$$S_{yy}(t_1, f) \equiv \int_{-\infty}^{\infty} R_{yy}(t_1, t_1 + \tau) \exp(-j2\pi f\tau) d\tau \quad (5)$$

with $\tau \equiv t_2 - t_1$. The latter equation defines an instantaneous power spectral density that may be a function of time t_1 in the non-stationary case where $R_{yy}(t_1, t_1 + \tau)$ is a function of time. Hence the whole problem reduces to the calculation of the autocorrelation function of the complex envelope

Calculation of the Autocorrelation of $y(t)$

Assuming that $\phi_1 \equiv \phi(t_1)$ and $\phi_2 \equiv \phi(t_2)$ are jointly gaussian random variables, equation (3) is found to be a particular case of the joint characteristic function

$$\Phi_{\phi_1 \phi_2}(\omega_1, \omega_2) \equiv E\{\exp j(\omega_1 \phi_1 + \omega_2 \phi_2)\}. \quad (6)$$

From the explicit expression of $\Phi_{\phi_1 \phi_2}(\omega_1, \omega_2)$ it is easy to show³ that

$$R_{yy}(t_1, t_2) = \exp j[\eta_{\phi}(t_2) - \eta_{\phi}(t_1)] \times \exp -[1/2 D_{\xi}^{(1)}(t_1, t_2)] \quad (7)$$

where $\eta_{\phi}(t) \equiv E\{\phi(t)\}$ is the instantaneous expected value of the phase and where

$$D_{\xi}^{(1)}(t_1, t_2) \equiv E\{[\xi(t_2) - \xi(t_1)]^2\} \quad (8)$$

is the first order non-stationary structure function⁴ of the zero-mean phase

$$\xi(t) \equiv \phi(t) - \eta_{\phi}(t) \quad (9)$$

$\xi(t)$ is identical to the phase noise process $\phi(t)$ except that its mean value is always equal to zero even when $\phi(t)$ is not stationary. Besides, the first increment of the phase may be expressed as a function of the moving average of the reduced frequency $y(t) \equiv [d\phi(t)/dt] / (2\pi\nu_0)$ yielding

$$D_{\xi}^{(1)}(t_1, t_1 + \tau) = (2\pi\nu_0\tau)^2 E\{\xi^2(t_1, \tau)\} \quad (10)$$

and

$$\eta_{\phi}(t_2) - \eta_{\phi}(t_1) \equiv (2\pi\nu_0\tau) \eta_y(t_1, \tau) \quad (11)$$

where

$$\zeta(t) \equiv y(t) - \eta_y(t) \quad (12)$$

is the zero-mean reduced frequency and where

$$\zeta(t, \tau) \equiv 1/\tau \int_t^{t+\tau} \zeta(x) dx \quad (13)$$

$$\eta_y(t, \tau) \equiv 1/\tau \int_t^{t+\tau} \eta_y(x) dx \quad (14)$$

are the moving averages of the zero-mean reduced frequency and of the instantaneous mean value of the reduced frequency. From the foregoing analysis we conclude that the stationarity of the RF power spectrum is directly dependent upon the stationarity of the frequency fluctuations. This result is to be compared with the analysis of Rutman⁵ who expresses the power spectrum as a function of the mean quadratic value of the stationary frequency fluctuations. Hence the modelling of non-stationary frequency fluctuations is a prerequisite to the further investigation of the RF power spectrum under non-stationary conditions.

Modelling of the Frequency Fluctuations

The polynomial model of frequency fluctuations is well known⁶. According to this model the power spectral density of the reduced frequency fluctuations may be expressed as a polynomial :

$$S_{yy}^+(f) \equiv \sum_{\alpha=-2}^2 h_{\alpha} f^{\alpha}. \quad (15)$$

Each term of the polynomial is considered as an independent process and the range $[-2, 2]$ of the exponent α is considered to be sufficient for the coverage of most noise sources in oscillators. For $\alpha \geq 0$ the frequency fluctuations are stationary while for $\alpha < 0$ the fluctuations belong to the flicker class of non-stationary processes⁷. Flicker processes are characterized by non-integer values of α in the range $[-2, 0]$. They can be modelled by fractional integration of a white noise⁸. The fractional integrator of order $-\alpha/2$ has a transfer function $H(p) \equiv p^{\alpha/2}$ and can be considered as a generalization of the ordinary integrator ($\alpha = -2$). Its impulse response is given by⁹

$$h(t) = u(t)[1/\Gamma(-\alpha/2)] t^{-\alpha/2-1} \quad (16)$$

for $\alpha < 0$ where $u(t)$ is the step function and $\Gamma(x)$ the gamma function. This filter is not

stable, hence the output is not stationary for a stationary input³. Assuming that the white noise fed to the input of the fractional integrator has a mean value equal to 0 and an autocorrelation function equal to $\frac{1}{2}(2\pi)^{-\alpha} h_\alpha \delta(\tau)$, the zero-mean frequency process $\zeta(t)$ can be modelled by the output of the fractional integrator. If the white noise is applied to the input of the fractional integrator at $t=0$, it can be shown that for $t_1 \leq t_2$ and $\tau \equiv t_2 - t_1$ the instantaneous mean quadratic value of the moving average of $\zeta(t)$ is given by

$$E\{\zeta^2(t_1, \tau)\} = \frac{1}{2}(2\pi)^{-\alpha} h_\alpha \int_0^{t_1+\tau} h^2(t_1-x) dx \quad (17)$$

Properties of the Power Spectrum

The analytic computation of (17) involves an hypergeometric function that may be expanded into a power series. The final form of the structure function is

$$\begin{aligned} \frac{1}{2} D_\zeta^{(1)}(t_1, t_1+\tau) = & \frac{1}{4} (2\pi)^{2-\alpha} h_\alpha v_0^2 \times \\ & [a \tau^{-\alpha-1} \\ & + \sum_{l=1}^{\infty} b_l t_1^{-\alpha-l} \tau^{l+1}] \end{aligned} \quad (18)$$

where a and the b_l are coefficients dependent upon α . Equation (18) is valid for $\alpha < 0$ except for negative odd integer values of α .

$\alpha \geq 0$: Stationary Power Spectrum

If $\alpha \geq 0$ the frequency fluctuations are stationary and so is the autocorrelation function of the complex envelope. Moreover the mean value of the frequency fluctuations and its moving average are equal and constant; hence the latter is independent of τ . Therefore the power spectrum of the oscillator signal is stationary. It should be noted that the power spectrum is dependent upon the bandwidth of the frequency fluctuations as well as upon α because the bandwidth determines as well the mean quadratic value of the moving average of the frequency.

$-1 \leq \alpha < 0$: Quasi-Stationary Spectrum

In this range of α the time dependent terms in (18) dampen out if sufficient time has elapsed since the beginning of the process. Thus the structure function is asymptotically stationary. However the frequency process itself is not stationary and its expected value,

as well as the moving average of the latter, is a function of time since the associated fractional integrator is unstable and since its step response does not dampen out. After equation (7), we see that the instantaneous power spectrum, as determined by the structure function, is stationary but drifts with time because it is centered about the moving average of the instantaneous mean frequency.

$\alpha < -1$: Non-Stationary Power Spectrum

If $\alpha < -1$ then the structure function is not stationary anymore since one or more time dependent terms in (18) do not dampen out. If a long time has elapsed since the beginning of the process, the time dependent term $i=1$ dominates all the others. The latter varies with τ^2 . Hence the autocorrelation function of the complex envelope is a gaussian curve which narrows more and more as time goes by. Accordingly the power spectrum is gaussian too and widens with time. This result is to be compared with Woodward's theorem¹⁰ which states that the power spectrum of an FM modulated carrier becomes gaussian as the modulation index increases, if the low-pass modulation process is gaussian. We have here the same situation except that the modulation index increases with time because of the non-stationarity of frequency fluctuations.

Influence of the duration of observation

Under the hypothesis of a band-limited stationary white phase noise, the width of the gaussian power spectrum is known to be proportional to the root mean square value of phase fluctuations¹. On the other hand, the mean square value of phase fluctuations under non-stationary conditions is dependent upon the duration of observation. In the case of a random walk of phase (brownian motion), for example, the accumulated mean square value is directly proportional to the elapsed time³. Moreover the spectrum analyser processes the input signal in just the same way whether the input is stationary or not. Indeed the foregoing ideas suggest that an equivalent stationary process of phase fluctuations possibly exists that would produce the correct non-stationary power spectrum for a given duration of observation. Such an equivalent process should take into account the dependence of the mean square value of phase fluctuations upon the duration of observation. Fig. 1 shows the power spectral density of an equivalent stationary phase noise process based on the polynomial model. In the range $[f_1, f_2]$ of the Fourier frequency the power spectral density of phase fluctuations is assumed to follow the

conventional polynomial law. f_1 is the low cut-off frequency defined as a constant divided by the duration T_{obs} of observation while f_2 is the high cut-off frequency defined as the bandwidth of the phase fluctuations. A more realistic modelling should include the actual behaviour of the low-pass cut-off band, yet the sharp cut-off approximation we use yields good agreement with experiment. In the range $[0, f_1]$ of the Fourier frequency, the power spectral density is assumed to be constant and equal to the value of the polynomial power spectral density at $f=f_1$. In this way, a dependency of the mean square value of the phase upon f_1 is introduced

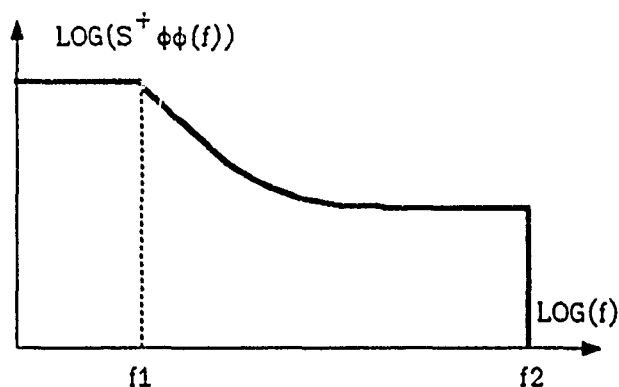


Fig 1 Power Spectral Density of the Equivalent Stationary Phase Noise Process versus the Fourier Frequency on log-log scales

Computer Calculation of the Power Spectrum

Considering the power spectral density from fig 1 as the definition of the equivalent stationary phase noise process, the associated autocorrelation function is the inverse Fourier transform of the power spectral density and can be computed analytically. Using this result, the autocorrelation function of the complex envelope may be expressed as a function of f_1 , f_2 and of the coefficients h_α of the polynomial model. Finally the power spectrum of the oscillator is obtained by numerical computation of the Fourier transform of the autocorrelation function of the complex envelope. The numerical integration applies the Gauss-Legendre method on discrete slices corresponding to half-periods of the oscillatory integrant. The non-linear Shank's algorithm¹¹ is used in order to accelerate the convergence of the resulting oscillatory sequence

Experimental Verification

Both the analytic non-stationary analysis and the numerical stationary analysis discussed in this paper were tested against experimental measurements for two particular cases. Experimental verification was conducted by frequency modulating a sinusoidal waveform generator by stochastic processes of known properties. The power spectrum of the output signal was measured using a digital FFT spectrum analyser. The final result is always the averaging of a number of individual spectra. In the case of non-stationary spectra each individual spectrum is centered before averaging.

$\alpha=0$ Low-Pass White Frequency Noise

For this first verification the generator was frequency modulated by a white noise source low-pass filtered by a first order filter. The analytical computation of the autocorrelation function of the complex envelope shows that the resulting power spectrum is bounded by two asymptotic cases: the gaussian and the lorentzian spectra¹², depending upon the ratio between the natural bandwidth:

$B_0 \equiv \pi/2 h_0 v_0^2$ of the frequency fluctuations and f_c , the -3dB cut-off frequency of the first order low-pass filter. h_0 is the white frequency noise coefficient from the polynomial model. If $B_0 < f_c$ then the power spectrum is

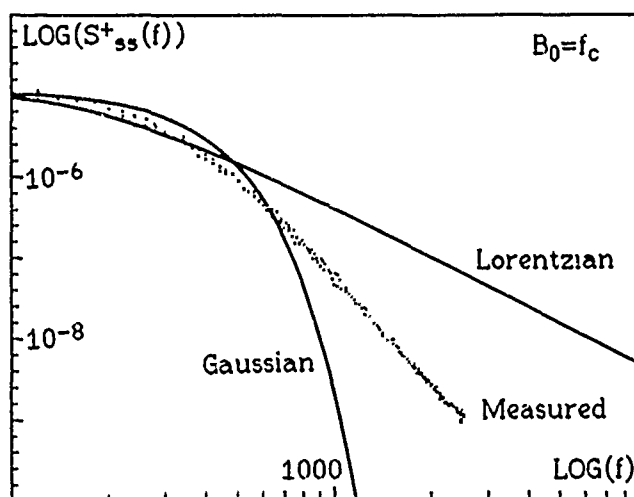


Fig. 2 Power Spectrum in $[V^2/Hz]$ versus the offset Fourier frequency in $[Hz]$ on log-log scales in the case of a band-limited white frequency noise

a lorentzian of half-power bandwidth B_0 . At Fourier frequencies much higher than the latter, the power spectrum follows a f^{-2} law and a measurement of $L(f)$ would yield the two-sided power spectral density of the phase noise, as expected from a low-noise stationary phase analysis. If $B_0 \gg f_c$, on the other hand, the power spectrum is a gaussian of bandwidth $4(\ln 2 B_0 f_c)^{1/2}$. The latter is proportional to the rms value of frequency fluctuations. This result is to be compared with the stationary phase situation where the bandwidth of the gaussian spectrum is proportional to the rms value of phase fluctuations. If $B_0 \approx f_c$ the power spectrum takes an intermediate form between the lorentzian and the gaussian asymptotic cases as shown on the experimental measurement of fig.2. Gaussian and lorentzian spectra can be produced by the computer calculation but are not shown here.

$\alpha = -2$ Random Walk of Frequency

For this second verification the generator was frequency modulated by a white noise source integrated once. The case $\alpha = -2$ corresponds to a non-stationary power spectrum and the finite duration of observation is taken into account by the introduction of a rectangular time window, of width T_{obs} , into the analytical computation of the autocorrelation function of the complex envelope. Defining the natural bandwidth $(B_{-2})^2 \equiv \frac{1}{2} T_{obs} (\pi \nu_0)^2 h_{-2}$, where h_{-2} is the coefficient of the polynomial model, the calculation yields that for

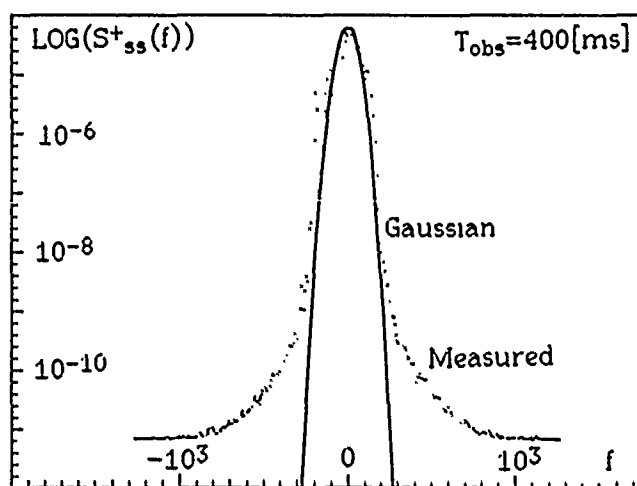


Fig.3 Power Spectrum in $[V^2/Hz]$ versus the offset Fourier frequency in $[Hz]$ on lin-log scales in the case of a random walk of frequency.

$B_{-2} \gg 1/T_{obs}$ the power spectrum is a gaussian curve of bandwidth $4[\ln 2 (B_{-2})^2]^{1/2}$. The latter is proportional to the root mean square value of the frequency fluctuations accumulated during the duration of observation. On fig.3 an experimental measurement of the power spectrum is compared against the gaussian theoretical curve for $T_{obs} = 400$ ms. It appears that for Fourier frequencies larger than the gaussian bandwidth, the experimental power spectrum follows a f^{-4} law which corresponds to the frequency dependence of the phase power spectral density. Indeed this behaviour of the experimental power spectrum is predicted by the computer calculation using the equivalent stationary phase process. Fig.4 shows the computed power spectrum as a function of the duration of observation

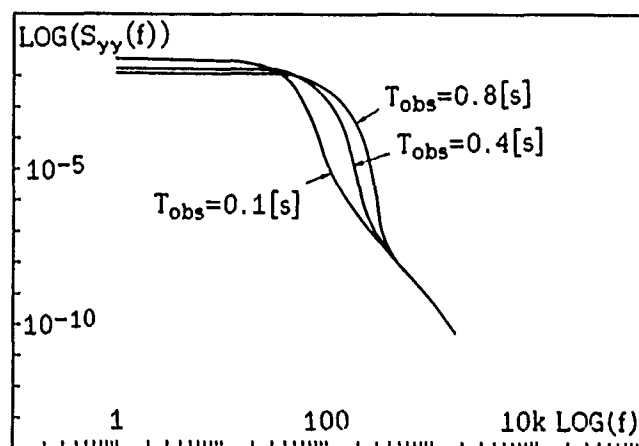


Fig.4 Power spectral density of the complex envelope in $[1/Hz]$ versus the Fourier frequency in $[Hz]$ on log-log scales in the case of a random walk of frequency as computed from the stationary equivalent phase noise process.

Conclusion

We have shown that the different terms of the polynomial model of frequency fluctuations yield three categories of power spectra depending upon the value of α . For $\alpha > 0$ the power spectrum is stationary. For $-1 \leq \alpha < 0$ the power spectrum is quasi-stationary, i.e. the instantaneous power spectrum is stationary but is centered about a drifting mean frequency. For $\alpha < -1$ the power spectrum is not stationary and its shape is dependent upon the duration of observation. Fractional integration

of a white noise process was used in order to model the non-stationary frequency fluctuations. This choice lead to the successful analytic calculation of the autocorrelation function of the complex envelope in the non-stationary case. Moreover an equivalent stationary phase noise process was defined and discussed, which enables a computer calculation of the power spectrum in the stationary case as well as in the non-stationary case. The calculation takes into account the effect of the duration of observation on the measured power spectrum in the non-stationary case. It was shown, finally, that for sufficiently short durations of observation a measurement of the non-stationary power spectrum may yield the two-sided power spectral density of phase fluctuations, as expected from the conventional low-noise stationary phase analysis. The non-stationary analysis presented in this paper was necessary, however, in order to determine under what conditions spectral purity and frequency stability may be identified one to the other and under what conditions they may not

References

- 1 L. G. Bernier, F. E. Gardiol, "The Analytic Signal Representation of Oscillators with Application to Frequency Stability Analysis," *Proc. of the 39th Annual Frequency Control Symposium*, Philadelphia, May 1985, pp.127-131
- 2 E. J. Baghdady, R. N. Lincoln, B. D. Nelin, "Short-Term Frequency Stability: Characterization, Theory and Measurement," *Proc. of the IEEE*, vol. 53, July 1965, pp. 704-722
- 3 A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill 1965.
- 4 W. C. Lindsey, C. M. Chie, "Theory of Oscillator Instability Based upon Structure Functions," *Proc. of the IEEE*, vol. 64, December 1976, pp. 1652-1666.
- 5 J. Rutman, "Instabilité de Fréquence des Oscillateurs," *L'Onde Electrique*, vol. 52, no. 11, décembre 1972, pp 480-487.
- 6 J. Rutman, "Characterization of Phase and Frequency Instabilities in Precision Frequency Sources: Fifteen Years of Progress," *Proc. of the IEEE*, vol. 66, no. 9, September 1978, pp. 1048-1075.
- 7 M. S. Keshner, "1/f Noise," *Proc. of the IEEE*, vol. 70, no. 3, March 1982, pp. 212-218
- 8 V. Radeka, "1/f Noise in Physical Measurements," *IEEE Trans. on Nuclear Science*, vol. NS-16, October 1969, pp. 17-35.
- 9 M. Abramowitz, I. A. Stegun, *Handbook of Mathematical Functions*, Dover, 1964.
- 10 N. M. Blachman, G. A. McAlpine, "The Spectrum of a High-Index FM Waveform: Woodward's Theorem Revisited," *IEEE Trans. on Communication Technology*, vol. COM-17, no. 2, April 1969, pp. 201-208.
- 11 R. W. Hamming, *Numerical Methods for Scientists and Engineers*, McGraw-Hill, 1973
- 12 A. Papoulis, "Random Modulation: A Review," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-31, no. 1, February 1983, pp. 96-105.

AGING MEASUREMENTS ON QUARTZ CRYSTALS IN THE BATCH MODE

E. Hafner and H. W. Jackson

XOTEX Corporation
P.O. Box 1026, Eatontown, NJ 07724Summary

A Quartz Crystal Measurement System has been developed that is capable of evaluating the full range of performance characteristics of crystal units, at any frequency from 1 MHz to 160 MHz. Its most outstanding feature is the ability to make concurrent aging measurements on up to 200 crystal units in a batch mode, with the resolution in average rate exceeding 1 pp 10E10/day for precision units. The system is equally capable of measuring, also in the batch mode, the frequency vs temperature and resistance vs temperature characteristics of crystals, within the range of the temperature chamber, as well as the drive level dependence of frequency and resistance with the drive level ranging from picowatts to milliwatts for most crystals. Standard parameters like C1 and Q, and the spurious mode spectrum, can also be evaluated with great precision, and these parameters can be evaluated at the intended operating temperature of the crystals, such as for ovenized oscillators. The system consists of two subsystems, a balanced bridge with associated electronics for the crystal parameter measurements and a temperature chamber with transport mechanism for sequencing the crystal units through the measuring head while in a highly controlled environment.

Keywords: Quartz Crystal Measurements, Crystal Aging, Temperature Chamber, Admittance Bridge.

Introduction

Aging measurements on quantities of quartz crystals pose an interesting combination of technical problems. Regardless of which of the various possible approaches is chosen, it is necessary that each crystal be measured repeatedly, over an extended period of time, under as nearly identical conditions as possible. In particular, the thermal and the electrical environment of the crystal at the time any one measurement is taken should be precisely the same as at any other measurement.

The only satisfactory techniques available for such measurements up till now employ banks of individual oscillators, coupled with suitable data acquisition systems that automatically measure, record and evaluate the frequency of each oscillator in turn, according to a predetermined schedule [1]. Aging measurements in the batch mode, where a larger number of crystal units is placed in a common temperature chamber, have apparently not been reported on for nearly two decades [2]. Such techniques are being employed, however, in a variety of implementations [3-6], primarily for the evaluation of the frequency-temperature characteristics of crystals, using a variety of methods for testing the crystals [7].

In this paper we report on the development of a new batch oriented aging measurement system that is capable of determining the average aging rates of even the highest precision crystal units. The aging behavior of the crystal resistance is evaluated as part of the frequency aging measuring routine. The system can equally be used for the high resolution measurement of all other electrical characteristics of the crystals in the temperature chamber.

The following is a brief description of this system, along with a discussion of some of the results obtained to date.

The Measurement SystemThe Bridge System.

The essential elements of the instrument used to measure crystal frequency and resistance are a synthesizer, an electronically tunable admittance bridge, a heterodyne receiver, a control unit and a desk top computer that interfaces with the control unit. A block diagram of the system is shown in Figure 1.

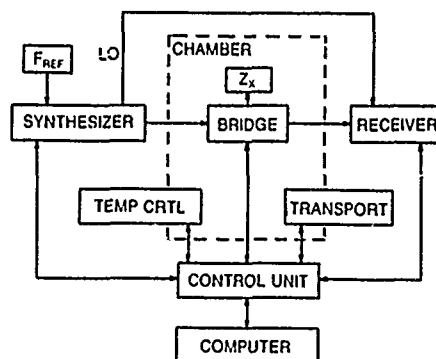


Fig. 1 Block diagram of the measuring system

The synthesizer covers the range from below 1 MHz to 160 MHz, with an effective resolution of 25 microhertz. The output signal level can be set, by means of a programmable attenuator, to any value from +10 dBm to below -100 dBm in 1 dB steps. The synthesizer also provides a second, tracking, signal that is offset from the main output by the receiver IF frequency, at any frequency within its range. This signal is fed to the receiver LO input. A Rubidium Standard was used as the reference for the synthesizer in the current work. The actual output frequency of the synthesizer is determined from the control word to

the synthesizer decades and from the output of a custom counter counting an internal signal frequency. It takes less than 100 milliseconds to read to a resolution of better than 1pp 10E10 at any frequency.

The Electronically Tunable Microcircuit Admittance Bridge is based on the design as described in the Proceedings of the 1976 and 1978 Frequency Control Symposia [8,9]. It is a Schering bridge with varactors as the variable capacitors. A simplified diagram is reproduced in Figure 2. In its current configuration the bridge has a frequency range of from 1 MHz to about 250 MHz, a resistance range from less than 3 Ohms to over 50 000 Ohms and a capacitance range of about ± 40 pF. The crystal leads can be clamped into the unknown terminals of the bridge, with a piano key clamp, at about two millimeters from the crystal base, either directly or through one of three load capacitors (20 pF, 30 pF and 100 pF). The voltages for tuning the bridge are derived from two 16 bit D/A converters, one for each of the two varactors. The bridge is calibrated by using microcircuit chip resistors and capacitors whose DC/low frequency values are accurately known. Routine calibration verification is accomplished by measuring the major and minor admittance components of standard 1% metal film resistors, mounted in crystal cans and carried along on the component transport, to be discussed below. Moreover, the initial balance of the bridge constitutes, essentially, a (single point) recalibration before each and every measurement of an unknown.

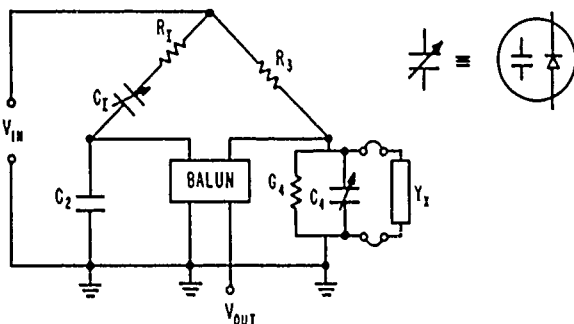


Fig. 2 Schematic of the balanced bridge network

The Receiver has a basic sensitivity of -130 dBm, software enhanced to -142 dBm. It employs single heterodyne conversion; its gain is computer controlled by way of external DC signals. Phase sensitive detection is not used; its function in the measurement process is replaced by software techniques. The LO signal is derived from the synthesizer as described above. The rectified output goes to an A/D converter.

The Control Unit interacts with the computer on one side, via an interface card, and the synthesizer, bridge and receiver on the other side. It also interacts with the temperature chamber and the component transport mechanism. It is designed to use only commercially available components. Several of the outputs of the control unit are high resolution DC signals, on long leads. Adequate precautions against common mode and thermal EMF problems were found to be essential to assure proper operation of the system.

The Computer currently used is a Zenith Z-100 desktop system with one hard and one floppy disk. The interface card to the control unit contains two 8 bit parallel ports. The software is compiled BASIC. It is

menu driven and provides for the automatic measurement of the selected performance aspects of any one crystal on the transport, either singly, or in rotation with all other crystals. Its speed in measuring the crystal frequency and resistance is limited mainly by the Q of the crystal under test. Repetitive measurements on a 5MHz 5th overtone crystal with a Q of 2 million take in the order of 15 seconds each, whereby each measurement consists of an initial balance off resonance and the final balance at resonance. 50 MHz crystals with a Q of 100 000 take about three seconds. The resolution of the bridge balance is limited by the applied drive level and the sensitivity of the receiver. The maximum available receiver sensitivity is, as stated above, -142 dBm.

The Temperature Chamber and Transport Mechanism.

The temperature chamber has an internal volume of 8 cubic feet. Its control range is $+45^{\circ}\text{C}$ to $+125^{\circ}\text{C}$; the control element is a platinum resistance thermometer. After stabilization for some three hours, the temperature in front of the bridge, where the crystals are measured, is constant to better than ± 20 millidegrees C. The bridge, with clamp arrangement, is mounted in about the center of the chamber by way of brackets extending from the rear of the chamber. The clamp is opened by means of a solenoid operated draw string and closed by spring action. Rf signals are fed to and from the bridge by semirigid coax cables. The chamber temperature is under computer control.

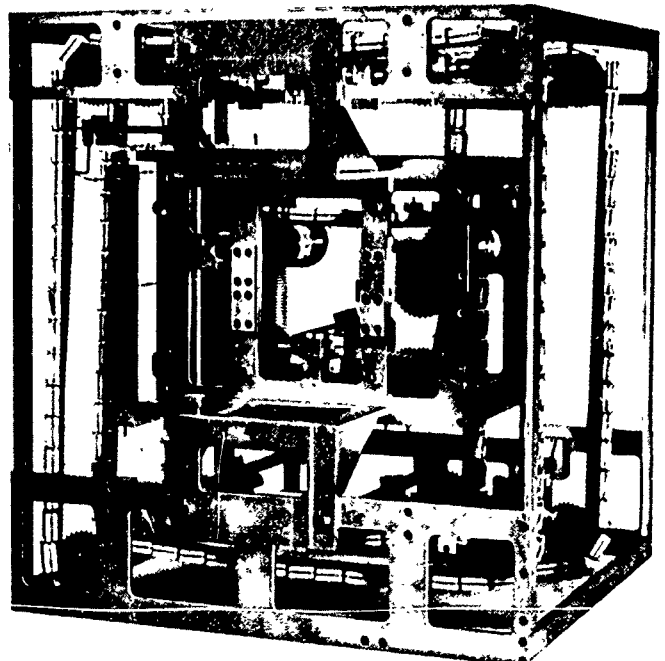


Fig. 3 Front view of transport rack

The transport rack, shown in Figure 3, is an open frame supporting a pair of parallel endless chains that run on sprocket wheels. The chains are about 1/2 inch apart, joined by brackets into which the individual crystal carriers can be inserted. There is room for more than 200 such carriers along the length of this chain drive. One pair of the sprocket wheels carrying the chain is motor driven by means of a drive shaft through the rear chamber wall. The run of the chain drive is arranged to pass in front of the bridge. The crystals, e.g. of the HC-6 or similar

type, are inserted into the carriers such that, when they are in front of the bridge with the clamp open, their pins extend over the bridge terminals, ready to be clamped in. The carriers for other type crystal enclosures, e.g. HC-40 or TO-5, have miniature PCB adapters from which the pins to be clamped into the bridge extend. These crystals are usually soldered into the PCB adapters. When in front of the bridge, the presence of the crystal pins is sensed by optical fibers, enabling the computer to position the pins relative to the bridge terminals to within a fraction of a millimeter. The entire transport rack with chain drive can be removed from the chamber to facilitate placement and removal of the crystal units and/or crystal carriers. Once in the chamber, the transport mechanism is operated under computer control. A window in the chamber door permits visual observation.

The Resolution of the Bridge Measurement.

During the initial balance of the bridge, the two internal capacitances shown in Figure 2, C_I and C_4 , are adjusted until the output of the bridge is zero, i. e., the bridge is balanced. If the unknown then inserted into the bridge terminals were a pure reactance, only C_4 would need to be changed to re-establish balance; if the unknown were a pure resistance, only C_I would need to be changed. The resolution of the bridge is determined by the amounts C_I or C_4 must deviate from their respective values at true balance, for the bridge output to exceed the effective noise level at the receiver input. Further, for most crystal measurements, C_4 is left at the value found during initial balance, and C_I and the synthesizer frequency are varied to balance the bridge [10]. Any uncertainty in the true value of C_4 will, therefore, cause a corresponding uncertainty in the crystal frequency. Moreover, even if the C_I and C_4 balance values are perfect, the finite receiver sensitivity will limit the degree to which the true resonance frequency can be determined.

Analytically it can be shown, by considering the input/output relationships of the bridge near balance, that

$$\Delta C_{4min} = (1/\alpha)(1/\omega)(P_{Dmin}/P_{sig})^{1/2} \quad (1)$$

$$\Delta C_{Imin} = (1/\alpha')(P_{Dmin}/P_{sig})^{1/2}, \quad (2)$$

where P_{Dmin} is the minimum detectable signal power at the receiver input and P_{sig} is the input power to the bridge. The bridge attenuation constant α is in the order of .08 for the microcircuit bridge, and α' is given by

$$\alpha' = (\alpha/R_3)(C_2/C_I)(1/C_I). \quad (3)$$

It can be shown further, from the relation for the minimum detectable phase angle of the bridge unknown [11]

$$\tan \phi_{min} = (R_x/\alpha)(P_{Dmin}/P_{sig})^{1/2}, \quad (4)$$

that the frequency error due to ΔC_4 is:

$$(\Delta f/f)_{C4} = (R_x/2Q)\omega \Delta C_4 \quad (5)$$

and due to the finite receiver sensitivity:

$$(\Delta f/f)_{min} = (1/\alpha)(R_x/2Q)(P_{Dmin}/P_{sig})^{1/2}. \quad (6)$$

In Eqs. (5) and (6), R_x is the resistance of the crystal and Q its quality factor. Approximate values for the constant elements in the microcircuit bridge are $R_I = 9$ Ohms, $R_3 = 10$ Ohms, $C_2 = 100$ pF. The voltage variable capacitors C_I and C_4 are approximately at 50 pF and 100 pF at initial balance, respectively. $R_{41} = 1/G_{41}$ is 5 Ohms.

The balance conditions for the bridge in Fig.3 are, neglecting parasitic elements,

$$G_4 + G_x = (C_2/R_3)(1/C_I) \quad (7a)$$

$$C_4 + C_x = (C_2/R_3)R_{41}. \quad (7b)$$

The values of the unknowns, G_x and C_x , are found from

$$G_x = (C_2/R_3)(1/C_{I1} - 1/C_{I0}) \quad (8a)$$

$$C_x = C_{40} - C_{41}, \quad (8b)$$

where the second subscripts 0 and 1 refer to initial balance, with $G_x = 0$ and $C_x = 0$, and to final balance respectively.

Among the other factors affecting the resolution of the bridge measurements are common mode signals on the bridge control lines, and spurious signals from the synthesizer - if their mixing products fall within the receiver bandwidth. Neither will be considered further here. The relations given above will be used later on when the measurement of drive level effects are discussed.

Measurements and Results

Batch Aging Measurements.

Up to 200 crystal units can be installed in the transport rack without regard to their mix in holder style, frequency or overtone. Operator interaction is required to input the frequency, load capacitance and drive level for each of the crystals loaded onto the chain into a file, which governs the automatic sequencing of the measurements from then on. With a fully loaded rack, about seven minutes are available for each crystal per 24 hours of aging time. In the normal mode of operation for aging measurements, each crystal in turn will remain clamped into the selected load terminal of the bridge for the full seven minutes each day, during which time a large number of measurements are being accumulated. Each measurement consists of an initial balance, off resonance, and a final balance at resonance. The data taken during the latter half of the measurement time on each crystal is averaged to yield a single value for frequency and resistance for this crystal on that day. The standard deviations from these frequency averages for any one day are in the low $10E10$ /high $10E11$ range, for well behaved high precision crystals and under normal operating conditions, e.g. crystal drive levels of 1/2 to 1 microwatt.

One day during the aging period is usually set aside for the F/T - R/T measurements. These are five point measurements covering a range of 15°C centered on the normal aging temperature, which is usually 80°C . With 200 crystals, and a stabilization time of two hours, about one minute is available to measure each crystal at each of four temperatures - 72.5 , 76.5 , 83.3 and 87.5°C . This is adequate even for the

highest Q crystals. The fifth point, at 80 °C, is taken from the preceding day's data. The coefficients for the F/T curve, and the turnover temperature, are subsequently determined by a least squares routine. The two hour stabilization time that we are currently using is certainly sufficient for high precision measurements. Much shorter times are likely to be adequate for most routine applications.

Figure 4 shows the long term aging behavior of a 10 MHz 3rd overtone SC crystal unit at 80.3 °C. This unit had remained on the transport for the first 87 days, was removed and placed on the shelf at room temperature for nearly two months and was then included again with an other batch of crystals for another 30 day aging run. During the 180 day period recorded in Figure 4 the chamber was opened and the rack removed on numerous occasions. The frequencies plotted are the raw data, with no adjustments made and no points removed. The vertical scale is in parts per 10E10. The average slope during the final 30 days is about 1.5 pp 10E11/day. The Figure indicates not only that Crystal No. SC11/3rd is a very stable unit, it also illustrates that the measuring system is, in fact, capable of evaluating such stable units.

Figure 5 shows the frequency aging of three 5 MHz 5th overtone AT's during the same thirty day run that is included as the final segment in Figure 4. The somewhat erratic nature of the traces, particularly during the early phases of the aging run, is seen in a large number of crystals whose aging rate is low enough for the curves to be plotted on the pp 10E10 scale. All three of the curves in Figure 5 show a step change on day nine into the test, which is coincident with the time the F/T - R/T test was performed, in the manner as described above.

Figure 6 shows the frequency and resistance of a 1 MHz fundamental mode AT crystal. The abscissa covers the same 180 day period as in Figure 4, (as does the abscissa in Figs. 6 and 7). The discontinuities in the curves coincide with the times the chamber was opened, briefly or for several days. Fig. 7 shows some sample aging curves from a batch of VHF 5th overtone AT crystals. In this run, the chamber was opened on day 26 into the test, which again caused a discontinuity in the curves. In fact, opening the chamber during an aging run appears to be a valid test for the susceptibility of the crystals to gross perturbations in their operating temperature.

The determination of the temperature stability of a crystal unit, while it is being measured for aging, is a difficult task that has not yet been accomplished satisfactorily. All indications are that, in our system, the stability is in the millidegree C range. Figure 8 is a plot of the temperature variations as seen by two different SC crystals, about 30 cm apart on the chain. The curves are the B mode frequencies of these crystals, measured in normal rotation with all other crystals and converted to temperature by way of their respective F/T coefficients (-30 ppm/°C). The two curves are reasonably well correlated, each indicating variations over the 30 day test period of less than +/- 10 millidegrees from the mean. From other measurements it is concluded that the upper bound for short term fluctuations is in the order of 1 millidegree per 180 seconds.

The effects of day to day temperature variations on crystal frequency are, undoubtedly, contributing to some extent to the scatter in the aging curves in Figs. 4 through 7. All attempts,

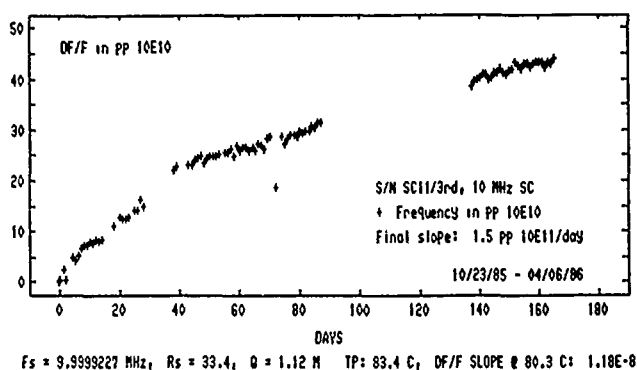


Fig. 4 Long term aging of a 10 MHz 3rd SC-cut

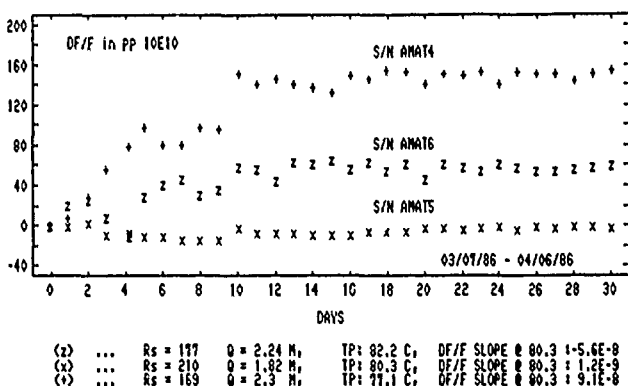


Fig. 5 Aging of 5 MHz 5th AT-cut crystals

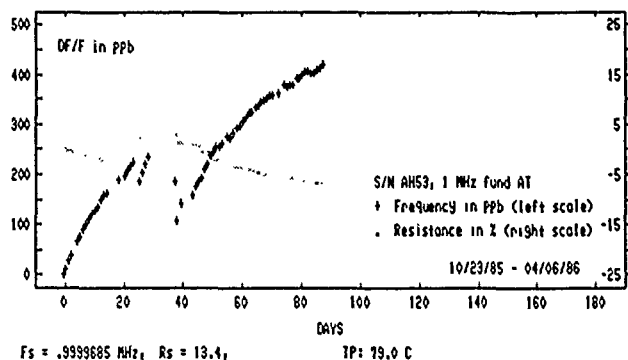


Fig. 6 Aging of a 1 MHz fund AT-cut crystal

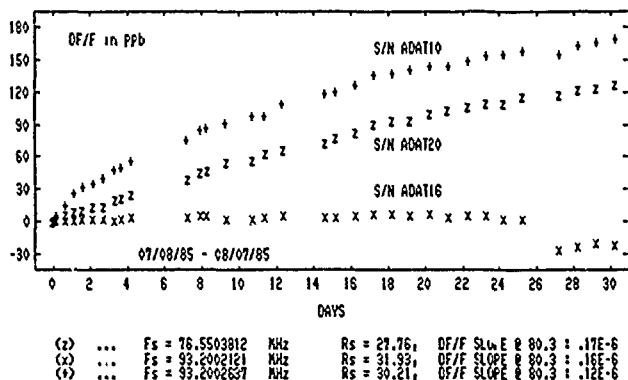


Fig. 7 Aging of VHF 5th AT-cut crystals

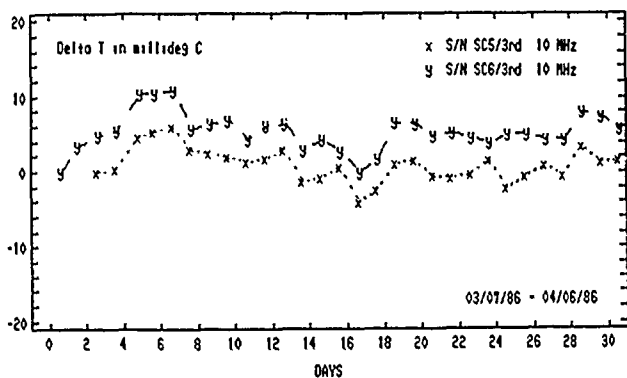


Fig. 8 Delta temp (in millideg C), from SC/3rd B modes

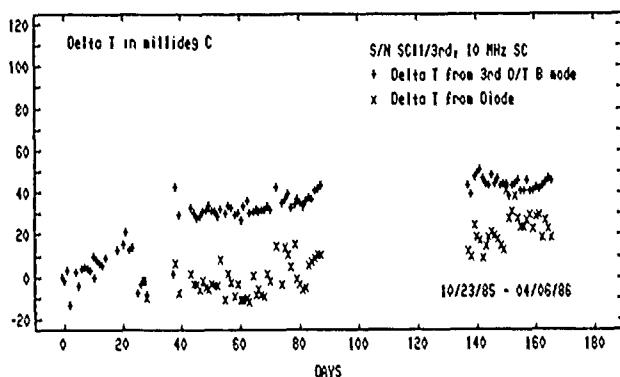


Fig. 9 Apparent temperature changes derived from SC/3rd B mode and from diode in bridge block

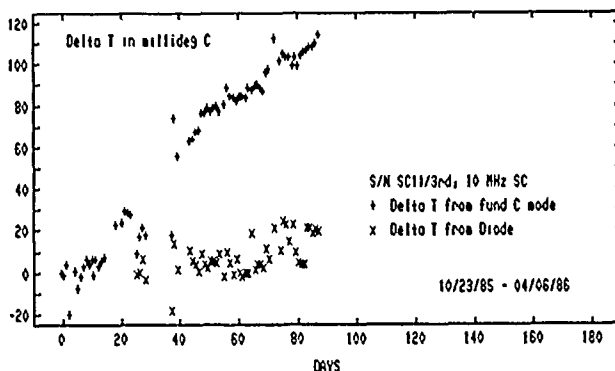


Fig. 10 Apparent temperature change per SC/fund C mode and per diode in bridge block

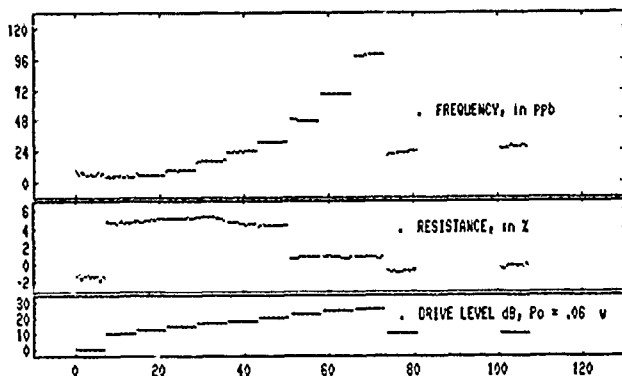


Fig. 11 Effects of drive level changes on a defective 5 MHz/5th AT crystal

however, to improve the smoothness of the curves by applying corrections based on the static F/T characteristics of the crystals have been unsuccessful to date, regardless of the thermometry used. Dynamic F/T effects [12,13] were not included in this effort. Until further evidence to the contrary becomes available we have concluded for the time being that, in our system, the temperature induced frequency fluctuations are a relatively minor part of the scatter in the aging curves for crystals near turnover. The larger part appears to be associated with the crystal units under test.

While Figure 8 gives the temperature variations apparently experienced by two different, well aged SC/3rd crystals over a 30 day period, Figure 9 shows a 180 day period for the same crystal as used in Fig. 4. Here again, the apparent temperature variations are derived from the third overtone B mode frequency ($T/C = -30 \text{ ppm}/^\circ\text{C}$). In a similar manner, Figure 10 shows the apparent temperature variations as derived from the fundamental C mode frequency ($T/C = -.8 \text{ ppm}/^\circ\text{C}$) of the same crystal, during the first 90 days of this test period. The fundamental B mode of this crystal gives a curve which is very similar to that for the 3rd O/T B mode. Included in Figs. 9 and 10 are the apparent temperature changes indicated by a diode inside the bridge block. (Occasional thermocouple readings, measurements with a resistance thermometer, as well as thermistor readings, indicated a constant temperature with a confidence of about ± 25 millidegrees). The fundamental C modes of other SC/3rd crystals carried on the chain indicated both, positive and negative apparent temperature changes, of greater and lesser magnitude than shown in Figure 10. We conclude that the apparent long term temperature changes derived from the high drift crystal frequencies in Figs. 9 and 10, in particular those derived from the fundamental mode frequency of a 3rd overtone crystal, are not real temperature changes, but reflections of the aging characteristics of these high drift modes. The long term drifts in the chamber temperature appear to be less than 5 millidegrees per month. With these assumptions we find an average aging rate of -1 ppb/day for the fundamental C mode of the same crystal that ages less than $2 \text{ pp } 10\text{E}11/\text{day}$ on the 3rd O/T C mode. In general we find that the aging characteristics of the fundamental C and the 3rd O/T C modes in the same crystal are unrelated, at least for the group of SC's included in this test. (We note, however, that several of the SC's reported on here are factory reject units).

Drive Level Effects on Frequency and Resistance.

Variations of crystal resistance under changing drive level conditions, at the operating temperature, have been found to be useful indicators of intrinsic problems that affect the performance of a crystal in an oscillator, but are not apparent from other tests, including a 30 day aging test. An example is shown in Figure 11 for the same 5 MHz/5th AT crystal that is represented by the bottom trace in Fig. 5. In Figure 11, the staircase function on the bottom indicates the drive level, which was changed from .06 microwatt to .6 microwatt, then in steps of 2 db to about 40 microwatt and, at about 75 minutes into the test, back down to the .6 microwatt level. The frequency in the corresponding time slots of constant drive is indicated by the substantially horizontal line segments that form a parabolic pattern in the top portion of the Figure. Each dot represents an individual measurement. It should be noted that the

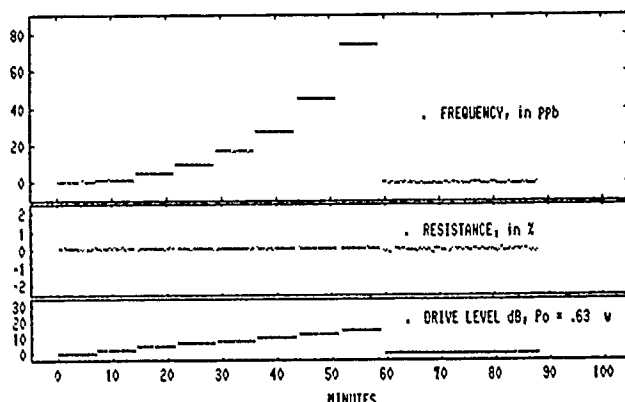


Fig. 12 Effects of drive level changes in an acceptable 5 MHz/5th AT crystal

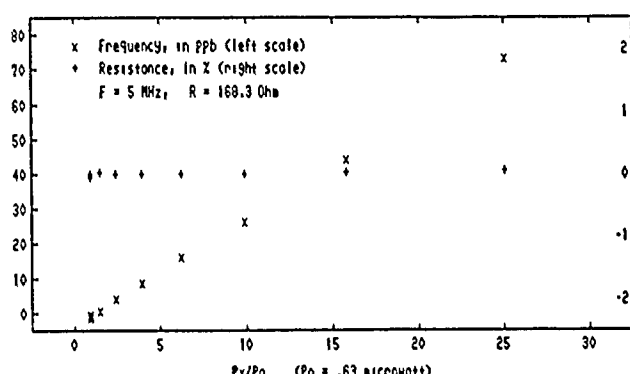


Fig. 13 Frequency and resistance vs drive from data in Fig. 12

frequency does not return to its initial value after the drive level is reset to the .6 microwatt level. The resistance, shown in the center section of the Figure, is seen to be generally erratic and to have several jumps with the largest in the order of 5%.

Figure 11 is to be compared with Figure 12, which shows the results of the same kind of measurement on a more typical, well behaved crystal, also a 5 MHz/5th AT unit. This is the crystal whose aging is plotted as the top curve, (+), in Fig. 5. As in Fig. 11, each dot of the plots for frequency and resistance in Fig. 12 represents an individual measurement consisting of an initial balance off resonance and a final balance at resonance. For these particular

tests, 24 such measurements were made at each drive level setting, except for the final setting, where four such groups of 24 measurements were taken. (In the test for Fig 11 no readings were taken during a 20 minute period between the 80 and 100 minute markers). It will be observed that the scatter in the data shown in Figure 12, particularly for resistance, decreases as the drive level increases, suggesting that the resolution is limited by thermal noise. Figure 13 is a restatement of the information in Fig. 12, with frequency and resistance now plotted as a function of the drive level. The frequency changes at a rate of about 2.9 ppb per microwatt. (For recent work on frequency - drive level effects see, e.g., [14])

Table I is a listing of the data from which Figure 13 was constructed, in part. The first column is the drive power ratio; the second is the frequency average, in pp 10E10, over the 24 individual readings shown in Figure 12 for any one drive level setting; SIGF is the square root of the Allen variance of the individual readings, also in pp 10E10. Rs (in Ohms) and SIGR (in percent) are the average resistance and its standard deviation at the various drive levels. C10 and C4, in picofarads, are the capacitance values for initial balance outside the resonance range of the crystal, but with the crystal in the bridge, and CI is the value for final balance at the crystal resonance; (C4 here is composed of the internal capacitance value of the bridge plus, essentially, the crystal Co in parallel). The sigma values for these capacitances, also in picofarad, are of particular interest, because they illustrate that the resolution of the measuring system is, in fact, near the thermal noise limit of the receiver.

With an effective receiver sensitivity of -142 dBm and an rf input into the bridge of -10 dBm (the level corresponding to the Po value used in Figures 12 and 13, and in Table I) we calculate from Eqs. (1) and (2) that the minimum detectable change in the capacitances is .7 femtofarad for CI and C10, and .09 picofarad for C4, at 5 MHz. The minimum detectable changes should become smaller with the inverse square root of the signal power applied to the bridge. These numbers are consistent with our results as reflected in the SIGC10, SIGC1 and SIGC4 values listed in Table I. We calculate further, from Eq. (5), that an error of .1 picofarad in the true value of C4 will cause an error of 1.2 pp 10E10 in the measurement of the resonance frequency, if the crystal has a Q of 2 million and a resistance of 150 Ohms. (The same value is obtained from Eq. (6)). This, too, is reasonably consistent with the observed SIGF values listed in Table I

(Smaller SIGF values are expected, and in fact observed at times, for lower resistance crystals).

TABLE I

DRIVELEVEL EFFECT ON AT 5 MHZ/5TH, S/N AMAT4, 4/16/1986

Px/Po	DF/F	SIGF	Rs	SIGR	C10	SIGC10	CI	SIGCI	C4	SIGC4
1.00	-123.5	1.7	168.290	0.05	50.2462	0.0006	48.8054	0.0005	96.6152	0.0738
1.58	-115.8	1.9	168.320	0.04	50.2454	0.0005	48.8048	0.0002	95.9007	0.0641
2.51	-80.8	1.9	168.280	0.02	50.2441	0.0004	48.8033	0.0003	95.9361	0.0586
3.98	-33.1	1.6	168.290	0.02	50.2422	0.0003	48.8016	0.0001	95.9875	0.0289
6.31	39.1	1.2	168.300	0.02	50.2397	0.0002	48.7994	0.0002	96.0290	0.0331
10.00	144.1	1.1	168.280	0.02	50.2362	0.0002	48.7960	0.0003	95.9139	0.0279
15.85	322.3	1.8	168.330	0.01	50.2303	0.0001	48.7907	0.0001	95.9764	0.0174
25.12	612.9	1.8	168.370	0.01	50.2213	0.0001	48.7826	0.0001	96.0241	0.0150
1.00	-131.1	1.2	168.180	0.06	50.2452	0.0007	48.8036	0.0006	95.7333	0.0832
1.00	-131.4	1.4	168.200	0.07	50.2450	0.0007	48.8035	0.0008	95.7804	0.0707
1.00	-131.3	1.5	168.220	0.07	50.2451	0.0006	48.8038	0.0006	95.7804	0.0807
1.00	-131.8	1.2	168.190	0.08	50.2449	0.0009	48.8033	0.0006	95.7846	0.0582

Frequency and resistance measurements as a function of drive have been made on numerous crystals with drive levels as low as 20 picowatt and as high as several hundred microwatt, to the milliwatt range on some. The resistance of a well behaved crystal is found to be essentially independent of drive, to within the limits of the measurement resolution.

The resolution of the bridge measurement can, of course, not be equated with its precision. Factors such as spurious synthesizer signals,

such as spurious synthesizer signals, environmental effects on the bridge and the associated electronics, etc., have to be considered before a definitive value can be assigned to its precision. However, the results on the aging tests presented above, and those of numerous other tests performed, indicate that the overall precision of the bridge measurements should be well within a factor of ten of its resolution. Its accuracy could, depending on the effort expended in its calibration, be made comparable to its precision. Difficulties in properly assessing the residual parasitic elements in the bridge, and their frequency dependence, are expected to impose practical limits on such efforts, however. Accuracies in the 1% range for resistance and 2% for capacitance appear attainable. The accuracy of resonance frequency measurements is a function of how closely the zero phase condition can be established during initial balance. It thus depends primarily on the precision of the bridge and is independent of its calibration. On the side of the crystal, the resonance frequency accuracy is prominently affected by lead length uncertainties, which cause an error that increases with the value of C_1 of the crystal and the operating frequency squared.

Conclusions

A new instrument has been developed for the batch measurement of quartz crystal units. Its performance exceeds the reported capabilities of previously available instrumentation. It is now possible, using the new measurement system, to evaluate the aging of up to 200 crystal units in the same temperature chamber with a resolution of better than 1pp 10E10/day for frequency and better than 1%/month for resistance. Beyond aging, any and all of the electrical characteristics of crystals are accessible to measurement by this system, and in nearly all cases the measurements can be performed automatically on all of the crystals in the batch. The high resolution measurement capability now established is expected to become especially valuable for oscillator engineers in need of an effective tool for screening crystal units before they are incorporated into the oscillators. It is suitable for laboratory as well as production applications.

Acknowledgement

The authors take pleasure in acknowledging gratefully the generosity of the Hewlett Packard Company, in particular of Mr. John Kusters, in providing samples of crystal units for experimentation during the early phases of this work.

References

- [1] Beetley, D. A., et al. "The quartz resonator automatic measurement facility", Proc. 35th Annu. Frequ. Control Symp., 1981, pp 263-270

- [2] Bernstein, M., "Precision measurement of the frequency aging of quartz crystal units", Proc. 22nd Annu. Frequ. Control Symp., 1968, pp 232-247
- [3] Pustarfy, H. S. and Smith, W. L., "An automatic crystal measuring system", Proc. 27th Annu. Frequ. Control Symp., 1973, pp 63-72
- [4] Metcalf, W. S., "Practical crystal measurements and standardisation", Proc. 27th Annu. Frequ. Control Symp., 1973, pp 55-58
- [5] Block, M. B., et al., "Results of temperature slewing quartz crystals for anomalous responses", Proc. 32nd Annu. Frequ. Control Symp., 1978, pp 344-353
- [6] Beaver, W. D., et al., "Quartz crystal measurements by a phase - amplitude method", Proc. 33rd Annu. Frequ. Control Symp., 1979, pp 189-200
- [7] Smythe, R. C. and Horton, W. H., "Evaluation of crystal measurement systems", Proc. 37th Annu. Frequ. Control Symp., 1983, pp 290-296
- [8] Hafner, E. and Riley, W., "Implementation of bridge measurement techniques for quartz crystal units", Proc. 30th Annu. Frequ. Control Symp., 1976, pp 92-102
- [9] Malinowski, G. and Hafner, E., "Automatic microcircuit bridge for measurements on quartz crystal units", Proc. 32nd Annu. Frequ. Control Symp., 1978, pp 354-364
- [10] Hafner, E., "The piezoelectric crystal unit - definitions and methods of measurement", Proc. IEEE 57, Feb. 1969, pp 179-201
- [11] Hafner, E., "Resonator and device measurements", Precision Frequency Control (E. A. Gerber and A. Ballato, eds.) Vol. 2, pp 1-44, Academic Press, Orlando, Florida, 1985
- [12] Ballato, A. and Vig, J. R., "Static and dynamic frequency - temperature behavior of singly and doubly rotated, oven - controlled quartz resonators", Proc. 32nd Annu. Frequ. Control Symp., 1980, pp 180-188
- [13] Stevens, D. S. and Tiersten, H. F., "Transient thermally induced frequency excursions in AT- and SC-cut quartz trapped energy resonators", Proc. 37th Annu. Frequ. Control Symp., 1983, pp 208-217
- [14] Tiersten, H. F. and Stevens, D. B., "The evaluation of the coefficient of nonlinear resonance for SC-cut quartz resonators", Proc. 39th Annu. Frequ. Control Symp., 1985, pp 325-332

THE PRECISE DETERMINATION BY AN AUTOMATIC SYSTEM ON THE RESONANCE
FREQUENCIES OF THE QUARTZ CRYSTAL RESONATOR.

Yoshimasa Oomura, Yasuaki Watanabe

Tokyo Metropolitan University, Faculty of Technology
2-1-1, Fukazawa Setagaya-ku, Tokyo 158, JAPANSummary

The Drop Method is a skillful method which resonance frequencies of a quartz crystal unit are more precisely obtained than another method. It is, however, so troublesome for us that the method has been scarcely used practically except research. In this paper it is presented that the automatic system was composed of Drop Method and applied to the measurement of 125MHz crystal units. The experimental results agreed well with the results by the manual measurement. The reproducibility of the measured resonance frequencies could be obtained in the range from 1.6×10^{-10} to 2.4×10^{-8} , the time for measurement was much more reduced than manual one.

1. Introduction

The resonance frequencies and the parameters of a crystal unit have been measured with several kinds of measuring method. They are, for example, CI meter, Pi-network, Fr-meter, Fl-meter, Drop Method and so on. Recently the network analyzer has been applied in U.S.A.¹⁾

On the other hand, an automatic measuring system, which is composed of a computer and equipments including the method's fixture mentioned above, has come into practical use. Especially the Pi-network has been supported by IEC²⁾, and also the Fr-meter³⁾ adopted in JIS (i.e. Japanese Industrial Standard). However it is necessary for them that the initial phase adjustment between input and output should be taken zero with the non-reactive resistor. After this preparation, the crystal unit is replaced instead of the resistor. Their measuring principle is so called substitution method.

The higher frequency stability of a quartz crystal oscillator is demanded in the fields of the contemporary telecommunication and the time keeping, the more precisely the resonance frequency of a crystal unit should be measured.

It seems, therefore, difficult for the substitution method to correspond to the higher stable measurement. The Drop Method⁴⁾ developed by the late Dr. I. Koga differs from the substitution method, and it is the skillful instrument that both the parameters and the resonance frequencies (i.e. f_s at series resonance, f_p at parallel resonance) of the crystal unit are more precisely obtained by the relations between frequency and voltage without depending on the phase measurement than another method mentioned above. Its measuring procedure, however, is so troublesome for us that the instrument has been scarcely used practically except research.

In this paper it is presented that an automatic system, which is composed of Drop Method's fixture, a personal computer and instruments, has been successfully developed. The experimental results obtained by the system on the 125MHz crystal unit agreed well with them by the manual measurement.

2. Preliminaries on measurement

The Drop Method's fixture and its equivalent circuit are shown in Fig.2-1 and Fig.2-2 respectively.

Fig.2-3 is an example of relation between impedance and exciting frequency of the crystal unit. It is understood from Fig.2-3 that two frequencies correspond to an impedance, and, therefore, the two loci are obtained by tracing the centers between two frequencies in the vicinity of the minimum impedance and the maximum one respectively. The loci were named by Dr. I. Koga the Mid Frequency Locus (abbreviated as MFL in the following). It is necessary for the Drop Method that the two procedures, that is, the crystal impedance measurement for the MFL and the processing of experimental result, are carried out.

Preparation :

The connecting piece is inset between terminals A and B as shown in Fig.2-2. The variable capacitor C_d is adjusted so that the circuit C_d - L_d - R_d may become non-reactive at the nominal resonance frequency of the crystal unit. After adjustment, the ratio (R_d/X_d) of the resistance R_d contained in L_d to the reactance X_d of C_d is obtained by the next relation; $V_0(2)$ and V_1 can be measured.

$$\frac{R_d}{X_d} = \frac{V_0(2)}{V_1} \quad (1)$$

And then the connecting piece is exchanged for the non-reactive resistor RRU⁵⁾ (Reference Resistor for everyday Use) of which resistance value is W . In this condition, the value of X_d can be obtained by the next relation.

$$X_d = \frac{W}{V_0(1)/V_1 - V_0(2)/V_1} \quad (2)$$

After decision of the X_d , the RRU is exchanged for the crystal unit under test. The relation which is composed of the crystal impedance Z_5 , R_d and X_d is denoted by the absolute value of the ratio of $V_0(1)$ to V_1 as shown in the next relation.

$$\frac{|Z_5 + R_d|^2}{X_d^2} = \left| \frac{V_0(1)}{V_1} \right|^2 \quad (3)$$

The Z_5 contains a new resistor element R_5 which is connected in series to the conventional equivalent circuit of the crystal unit. It is mentioned in appendix A-1) why the R_5 is adopted in this paper.

3. Measurement procedure by Drop Method system

It is necessary for the Drop Method that the two frequencies are measured at same impedance for making the MFL, and all the while the crystal current must be kept constant. In the case of an automatic measuring, the characteristic between frequency and impedance of a crystal unit under test is unknown, therefore, the method is put into practice as follows by considering the high speed measurement and the reproducibility of the obtained result. The obtained results shall be expressed by only around f_s without special explanation, because the impedance curvatures

to frequency are almost similar in the vicinity of both f_s and f_p .

3-1). Decision of frequency range:

After the impedance of a crystal unit is roughly measured in the vicinity of the nominal frequency, the frequency range from f_s to f_p is generally decided as shown in Fig.3-1. In the frequency range, the measurement is carried out again by maintaining the crystal current constant. The detailed measurement must be put in practice especially around f_s and f_p as shown in Fig. 3-2.

3-2). Approximation by polynomial expression:

By applying a high order polynomial to the obtained values in the item 3-1), two frequencies corresponding to an impedance, the frequencies (f_{min} and f_{max}) to minimum impedance and maximum one, and also these two impedance's values are obtained respectively as shown in Fig.3-3.

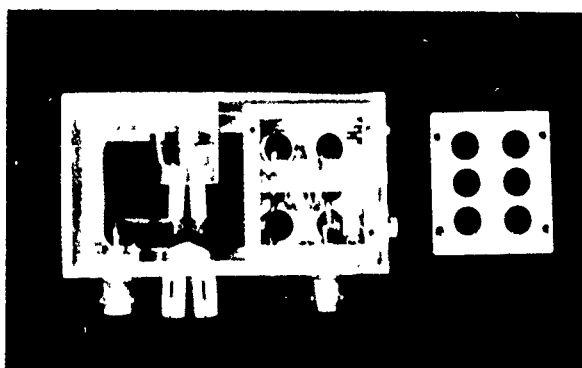


Fig.2-1 Picture of Drop Method fixture.

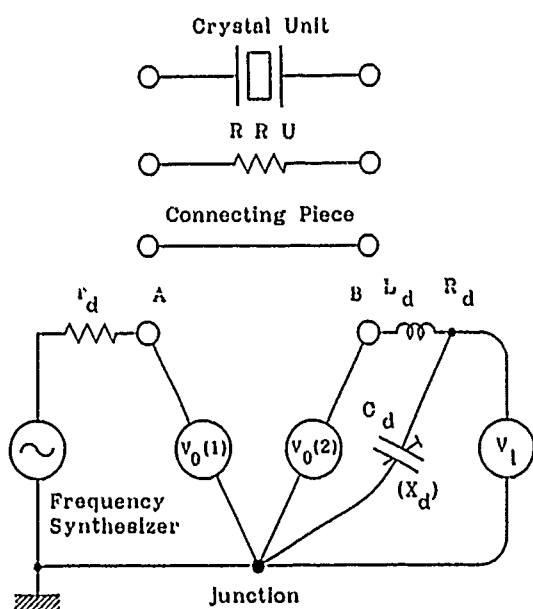


Fig.2-2 Equivalent circuit of Drop Method fixture.

3-3). Proper measurement:

The measurement is carried out concerning a pair of two frequencies obtained by the calculation in item 3-2). After the one (f_1) of two frequencies is fixed, the impedance (Z_1) at f_1 is decided by keeping the crystal current constant. The other frequency (f'_1) is so decided that the impedance (Z'_1) becomes the same value as Z_1 by changing frequency gradually up and down as shown in Fig.3-4. Consequently a pair of two frequencies can be decided at the same impedance, the MFL can be obtained by tracing the centers between two frequencies. A high order polynomial is applied to two MFLs in the vicinity of f_s and f_p respectively by using obtained values. The polynomial expression for MFL around f_s is shown in Fig.3-5. In this figure the impedance is denoted by the logarithm of absolute ratio squared of $V_0(1)$ to V_1 , and this valuable transformation is carried out by the relation of (A-1) as shown in appendix A-2).

3-4). Differentiation of MFL:

The applied polynomials to two MFLs are differentiated, and the differential coefficient of MFL on the part of f_s is calculated at some frequency. The frequency in f_p side is so decided that the absolute value of f_p side's differential coefficient becomes equal to the absolute value of obtained result in f_s side.

3-5). Decision of X_p/X_d :

The value of X_p/X_d can be decided by sum of two impedances at frequencies corresponding to calculation of differential coefficients in 3-4). The X_p is a reactance of capacitor C_p consisted of electrodes, that is, $X_p = 1/(4\pi C_p)$. Moreover, the MFL in f_p side is revolved around at the center of the frequency for 180 degree and is superposed on the MFL in f_s side by shifting it. The superposed result is printed out on plotter as shown in Fig.3-6.

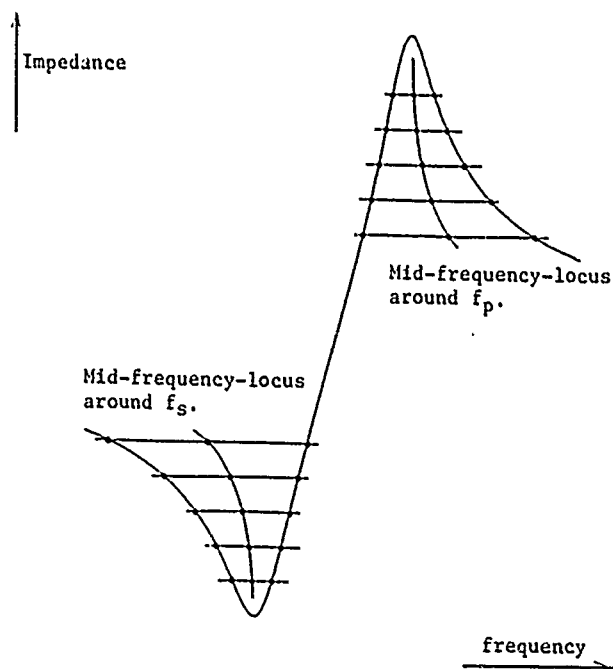


Fig.2-3 Mid-Frequency-Loci.

3-6). Decision of f_s and f_p :

After the variables are transformed by X_p/X_d , the two tangents corresponding to two MFLs are drawn as shown in Fig.3-7. The f_s and f_p are decided by the points of intersection between the tangents of MFL and the abscissa (frequency) respectively. Formulae applied here are shown by (A-2) and (A-3) in appendix A-2).

3-7). Decision of parameters in equivalent circuit:

The values of X_s , X_p , R_s and R_p can be obtained by the values of X_p/X_d and others already decided. The X_s is a reactance of the motional arm, that is, $X_s = \omega_s L_s = 1/(\omega_s C_s)$, and also R_s is a resistance included in motional arm. These parameters are obtained by the relations of (A-4) to (A-6) as shown in appendix A-2). The obtained resonance frequencies and parameters values are written on the printer, and stored in the memory. Fig.3-8 is the flow chart of an automatic system's program from the measurement to the processing of obtained results.

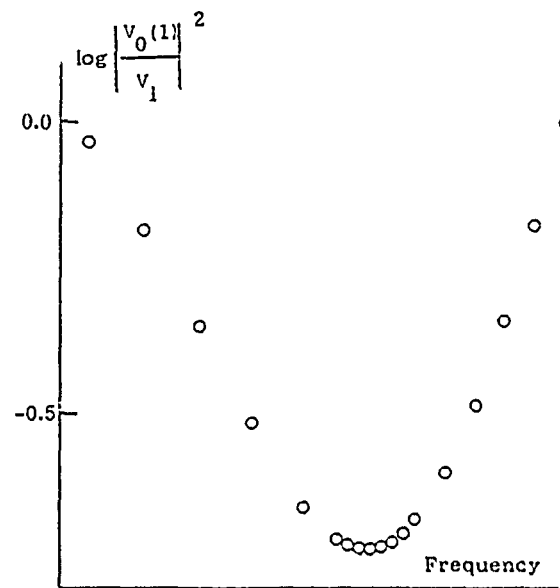


Fig.3-2 Detailed measurement in the vicinity of f_s .
(Same measurement in f_p side)

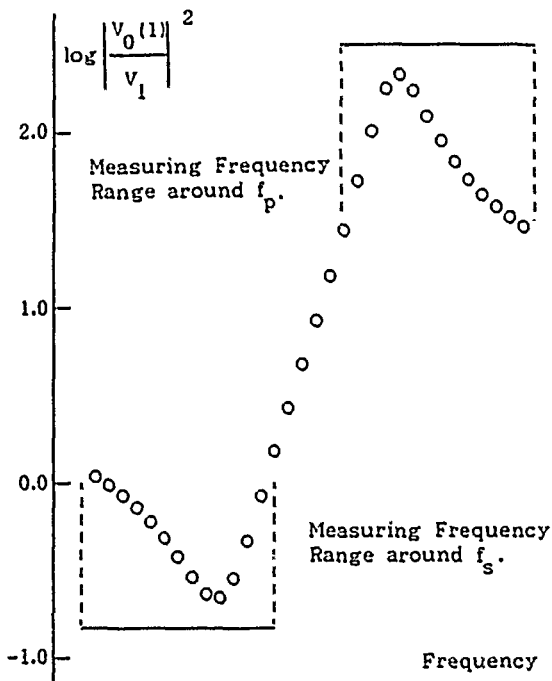


Fig.3-1 Decision of measurement range.

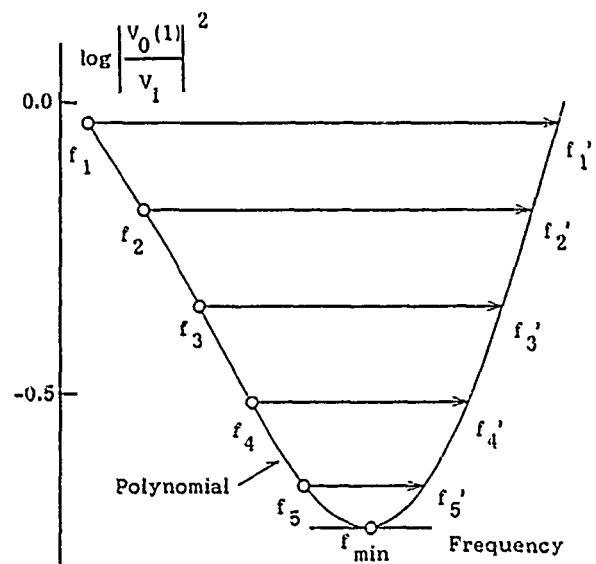


Fig.3-3 Polynomial expression for measured value in the vicinity of f_s .

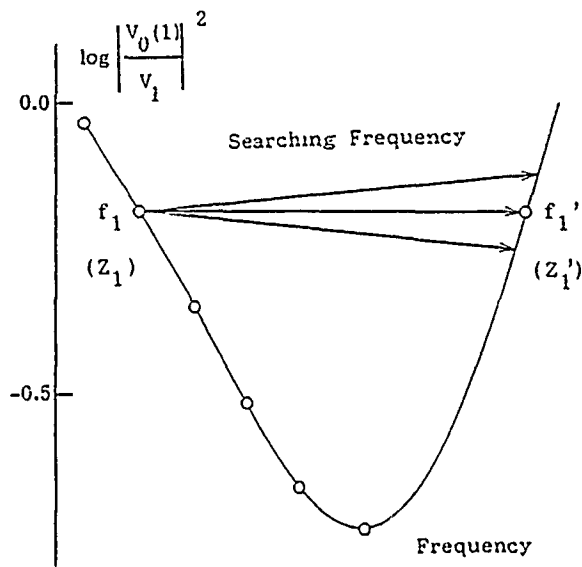


Fig. 3-4 Proper measurement in the vicinity of f_s .

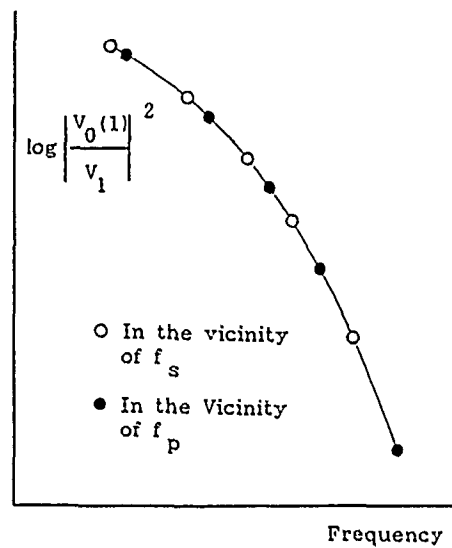


Fig. 3-6 Superposing Mid-Frequency-Loci. (Plotter output)

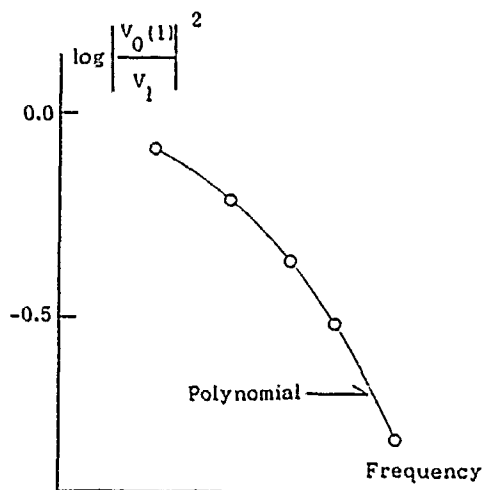


Fig. 3-5 Polynomial expression for Mid-Frequency-Locus in the vicinity of f_s .

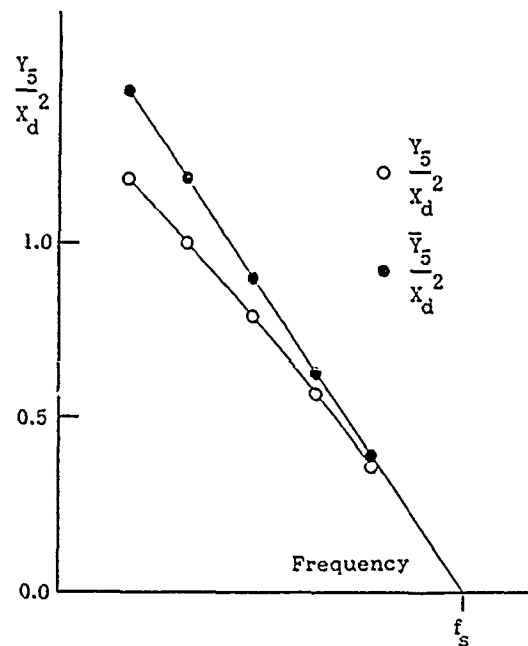


Fig. 3-7 Decision of f_s .

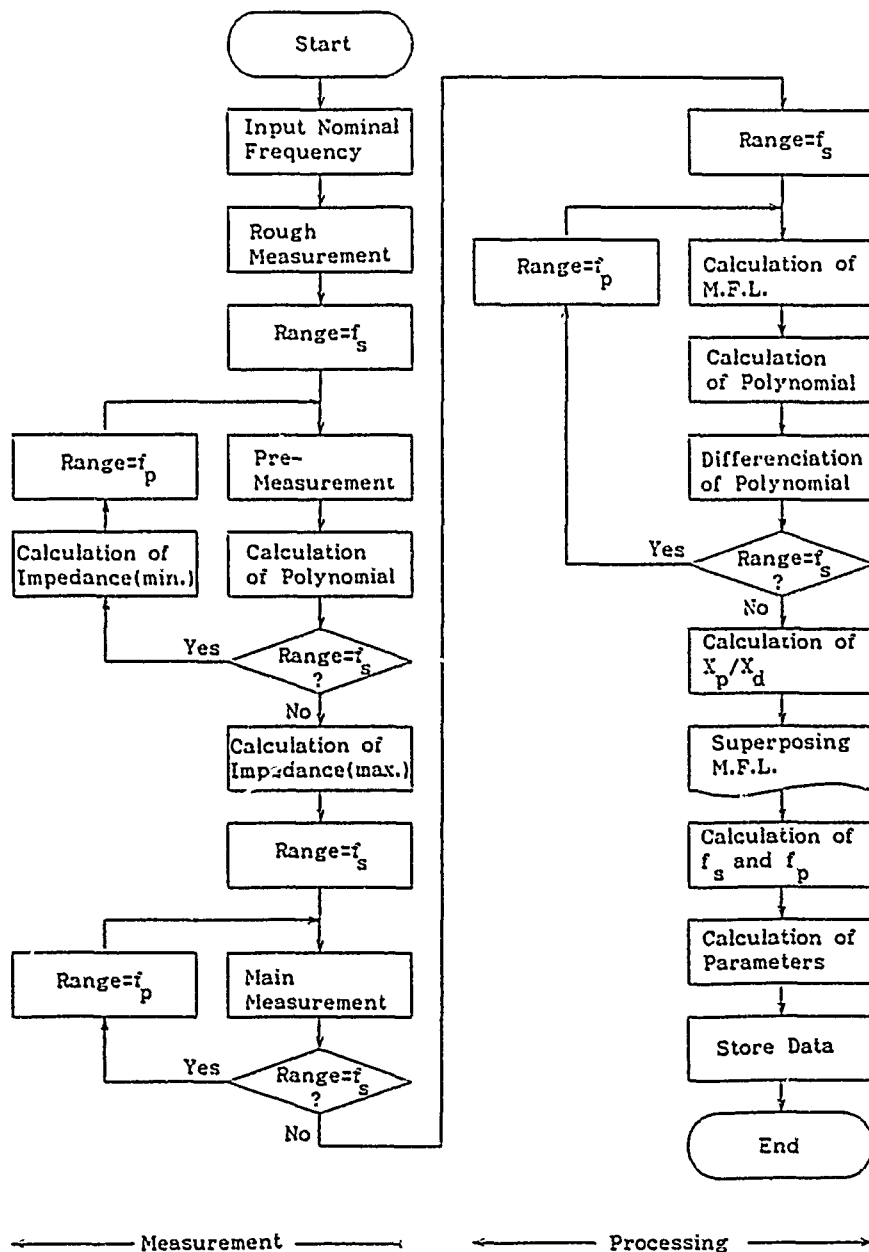


Fig.3-8 Flow-chart for computer program.

4. Experimental results

4-1). Measurement system and environment:

Fig.4-1 is the block diagram of an automatic measurement system. The measurement system including an oven is shown in Fig.4-2. Before starting, the rubidium atomic oscillator was prepared so as to keep the frequency synthesizer precise during measurement. Moreover, since the room temperature was kept $25 \pm 2^\circ\text{C}$ by an air conditioner, the temperature change in the oven including a crystal unit was able to be kept within $25 \pm 0.01^\circ\text{C}$. The electric power source of both frequency synthesizer and voltmeter had always been supplied previously to stabilize the system. Another instruments had been turned on for five hours before the measurement began.

4-2). Characteristics between frequency and current of crystal units under test:

Four crystal units of which nominal frequencies are 125MHz were selected as samples under test according to the relations between non-reactive frequency f_r , non-reactive resistance R_r and current squared I^2 , and were named as Ser. 74-1009, Ser. 74-1011, Ser.74-1013, Ser.74-1019 respectively. Their relations are shown in Fig.4-3. It is understood from the same figure that the relations of f_r to I^2 show straight lines and the R_r does not change against crystal current squared. It can be, therefore, said for these crystal units to be fit for the drop method measurement. The crystal current was kept 0.5mA constant all the while automatic measurement. Moreover the 6th order polynomial and the 3rd order one were adopted as polynomial in 3-2) and 3-4) respectively.

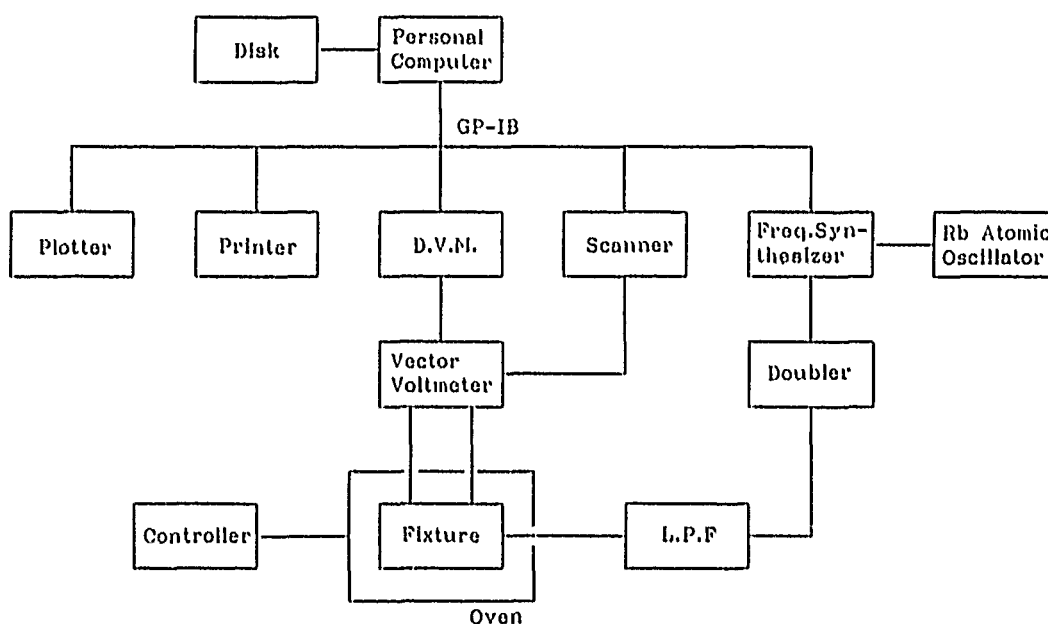


Fig.4-1 Block diagram of the automatic measuring system.

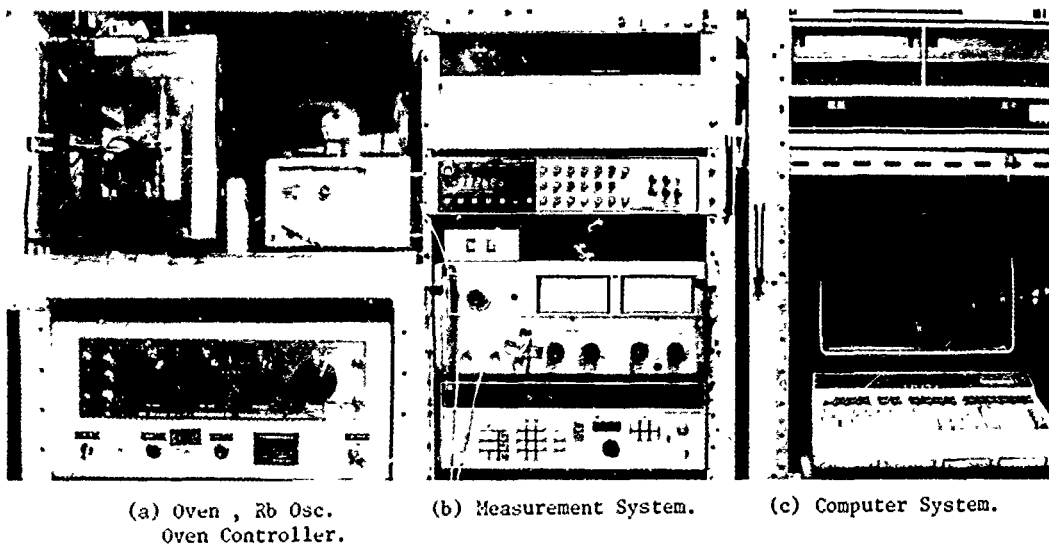


Fig.4-2 Picture of the system.

4-3). Comparison between results by automatic system and them by manual measurement:

Table-1 shows the measurement results by automatic system adopted in this paper and also by manual one. The data by manual are results which Dr. I. Koga had already measured about ten years ago. According to Table-1, it is obvious that the resonance frequencies f_s and f_p are agreed within -0.77ppm , the equivalent parameters are within $+9.8\%$ deviation. Even if the aging for ten years happened, it is considered that the crystal unit (Ser.74-1009) was well constructed. Therefore it is concluded that the good agreement was taken between them. Comparing the time required in measuring, however, it took 120 minutes approximately for the manual, on the other hand, about 10 minutes for the automatic system. Consequently the measuring time was reduced to almost $1/10$ time.

4-4). Reproducibility of measurement:

The reproducibility of measurement was confirmed by three times measurements at every one hour interval for each crystal unit under test. It is for the reasons of the procedure why the change of resonance frequencies and equivalent parameters' values must not be happened by putting on and off the crystal unit from fixture. Table-2 (a) to (d) shows the obtained results. It is understood from the table that the reproducibility is within $+2.4 \times 10^{-8}$ for the resonance frequencies, the best is -1.6×10^{-10} , and also $+3.9\%$ for the equivalent parameters.

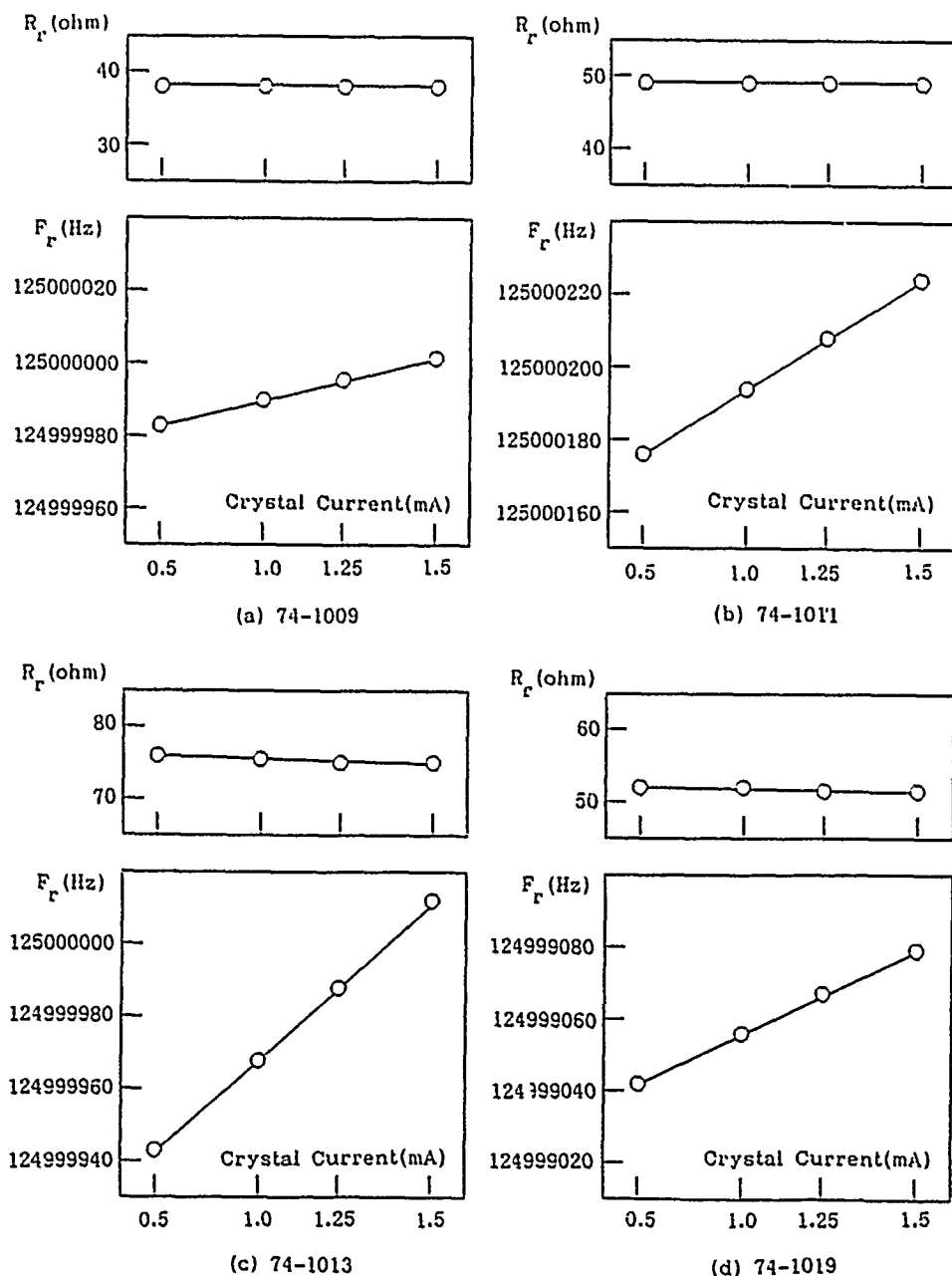


Fig.4-3 Current characteristics of quartz crystal units.

Table-1 , Measured Results.
(Quartz Unit Ser.74-1009 , $Q \approx 117000$, Temp.25°C)

	f_s (Hz)	f_p (Hz)	$R_s(\Omega)$	$R_p(\Omega)$	$X_p(\Omega)$	$X_s(\Omega)$
Manual Measurement	125000047	125004288	8.08	30.5	234	3.45×10^8
Automatic Measurement	124999951	125004226	8.80	27.5	221	3.23×10^8
Deviation	-0.77(ppm)	-0.50(ppm)	+8.9(%)	-9.8(%)	-5.6(%)	-6.7(%)

Table-2 , Measured Results.
(The number in parentheses is deviation from Number 1.)

(a) Quartz Unit Ser.74-1009 ($Q \approx 117000$, Temp.25°C)

Number	f_s (Hz)	f_p (Hz)	$R_s(\Omega)$	$R_p(\Omega)$	$X_p(\Omega)$	$X_s(\Omega)$
1	124999950.89	125004226.11	8.80	27.53	221.12	3.233×10^8
2	124999950.53 (-2.9 $\times 10^{-9}$)	125004224.40 (-1.4 $\times 10^{-8}$)	8.96 (1.8%)	27.40 (-0.5%)	220.77 (-0.2%)	3.228×10^8 (-0.2%)
3	124999950.87 (-1.6 $\times 10^{-10}$)	125004226.60 (3.9 $\times 10^{-9}$)	9.01 (2.4%)	27.33 (-0.7%)	220.20 (-0.4%)	3.219×10^8 (-0.4%)

(b) Quartz Unit Ser.74-1011 ($Q \approx 120000$, Temp.25°C)

Number	f_s (Hz)	f_p (Hz)	$R_s(\Omega)$	$R_p(\Omega)$	$X_p(\Omega)$	$X_s(\Omega)$
1	125000009.83	125003099.92	6.53	36.47	216.83	4.385×10^8
2	125000009.95 (9.6 $\times 10^{-10}$)	125003102.77 (2.3 $\times 10^{-8}$)	6.27 (-3.9%)	36.75 (0.7%)	217.82 (0.5%)	4.402×10^8 (0.4%)
3	125000009.87 (3.2 $\times 10^{-10}$)	125003100.79 (6.9 $\times 10^{-9}$)	6.59 (0.9%)	36.40 (-0.2%)	216.72 (-0.1%)	4.382×10^8 (-0.1%)

(c) Quartz Unit Ser.74-1013 ($Q \approx 113000$, Temp.25°C)

Number	f_s (Hz)	f_p (Hz)	$R_s(\Omega)$	$R_p(\Omega)$	$X_p(\Omega)$	$X_s(\Omega)$
1	124999729.62	125001602.63	6.89	61.02	207.29	6.917×10^8
2	124999729.07 (4.4 $\times 10^{-9}$)	125001601.33 (-1.1 $\times 10^{-8}$)	6.75 (-2.0%)	61.12 (-0.2%)	207.48 (-0.1%)	6.926×10^8 (-0.1%)
3	124999725.57 (-2.4 $\times 10^{-8}$)	125001603.32 (5.5 $\times 10^{-9}$)	6.93 (0.6%)	60.97 (-0.7%)	207.37 (-0.4%)	6.902×10^8 (-0.2%)

(d) Quartz Unit Ser.74-1019 ($Q \approx 115000$, Temp.25°C)

Number	f_s (Hz)	f_p (Hz)	$R_s(\Omega)$	$R_p(\Omega)$	$X_p(\Omega)$	$X_s(\Omega)$
1	124998927.15	125001843.56	7.89	41.16	220.70	4.730×10^8
2	124998925.37 (-1.4 $\times 10^{-8}$)	125001843.87 (2.5 $\times 10^{-9}$)	7.99 (1.3%)	41.07 (-0.2%)	220.44 (-0.1%)	4.721×10^8 (-0.2%)
3	124998926.86 (-2.3 $\times 10^{-9}$)	125001841.79 (-1.4 $\times 10^{-8}$)	8.10 (2.7%)	40.96 (-0.5%)	220.20 (-0.2%)	4.721×10^8 (-0.2%)

4-5). Confirmation of equivalent parameters:

The crystal unit's impedance behavior, which had been already obtained in the range from around f_s to f_p , is compared with the calculated impedance by using the five equivalent parameters obtained with the automatic system. In this case, formula (3) is rearranged by the relations of equivalent parameters as shown in next formula.

$$\left| \frac{Z_5 + R_d}{X_d^2} \right|^2 = \frac{X_p^2 \left[\alpha_s^2 + X_s^2 + 2\alpha_s (\alpha_5 + \beta) \right]}{X_d^2 \left[(1-X)^2 + \alpha_s^2 \right]} + (\alpha_5 + \beta)^2 \quad (4)$$

where $\alpha_s = \frac{R_s}{X_p}$, $\alpha_5 = \frac{R_5}{X_p}$, $\beta = \frac{R_d}{X_p}$, $X = \frac{\Delta f}{A} = \frac{f - f_s}{f_p - f_s}$

The result for Ser.74-1009 is shown in Fig.4-4. It is understood from Fig.4-4 that the calculated impedance curve fits well to the measured one. The same results as Fig.4-4 were obtained for another samples.

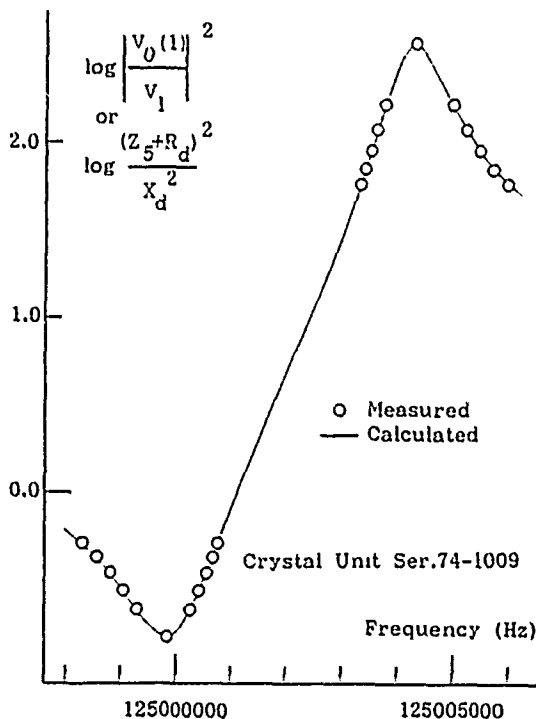


Fig.4-4 Comparison with measured and calculated value.

5. Conclusion

The automatic Drop Method system, which is composed of a computer, instruments and other, has been developed for the measurement of resonance frequencies and equivalent parameters of a crystal unit. The crystal units of 125MHz were practically measured with the automatic system. The obtained results agreed well with them by manual measurement. The reproducibility of measured values was better than $+2.4 \times 10^{-8}$ for the resonance frequencies (f_s, f_p), and $+3.9\%$ for the equivalent parameters. The measuring time was reduced to 1/10 time comparing with the time by the manual measurement.

Moreover, if the professional system is developed, the time will be more reduced than 10 minutes.

6. Acknowledgment

The authors wish to dedicate this paper to the memory of Dr. Issac Koga.

We would also like to thank Mr. S. Kobayashi, who is with KDD Laboratory, for preparation of crystal units.

7. Reference

- (1). EIA STANDARD "STANDARD METHOD FOR MEASUREMENT OF THE EQUIVALENT ELECTRICAL PARAMETERS OF QUARTZ CRYSTAL UNIT 1KHz to 1GHz." EIA-512, (April 1985).
- (2). IEC Publication 444: "Measurement of quartz crystal unit parameters by zero phase technique in a PI-network."
- (3). I.Koga: "Measurement of a VHF Crystal Unit Using an 'Fr-meter' and an 'Auxiliary Reference Resistor.'" Printed by University of Tokyo Press, (August 1980).
- (4). I.Koga: "The equivalent circuit of VHF crystal unit-equipped with evaporated electrodes." Published by Denki Tsushin Kyokai (Japanese), (April 1983). This book was translated into English for class by Y.Oomura and the note copies were distributed to EIA Conference held in April 1985 at Northern Illinois University.

Appendix

A-1). Reason of adopting R_5 :

The conventional equivalent circuit of a crystal unit is generally shown as Fig.A-1. If its impedance curve against exciting frequency is calculated according to the equivalent circuit, the obtained curve must have essentially a symmetric point. It was, however, found out by Dr. I.Koga that the impedance curve did not become symmetric exactly in the range f_m in the vicinity of f_s to f_p in accordance with the experimental results. Consequently, the calculated impedance curve by the equivalent circuit, which is consisted of four parameters, does not fit to the measured one in the case of a crystal unit equipped with evaporated electrodes. Moreover, it was confirmed that the difference of two impedances' sharpness between around f_s and f_p respectively became larger by adding a resistor in series to a crystal unit. It is natural, therefore, that a resistor R_5 should be contained primarily in the position as shown in Fig.A-2 for the reason of asymmetric impedance curve. According to the reference(4), it is explained that a resistor R_1 of motional arm in the conventional equivalent circuit must be divided into two resistors, that is, R_s and R_5 as shown in

Fig.A-2, and the R_5 will be caused by the viscosity between the evaporated electrode and the moving crystal plate at resonance. The new equivalent circuit is adopted in this paper for the above mentioned reason.

A-2). Formulae set up for obtaining parameters:

(1). The valuables transformation of MFL.

$$\left| \frac{V_0(1)}{V_1} \right|^2 = \frac{Y_5}{X_d^2} + \frac{(R_5 + R_d)^2}{X_d^2} \quad (A-1)$$

$$\begin{aligned} \text{where } Y_5 &= \left| Z_5 + R_d \right|^2 - (R_5 + R_d)^2 \\ &= X_p^2 \cdot \frac{\alpha_s^2 + X^2 + 2\alpha_s(\alpha_s + \beta)}{(1 - X)^2 + \alpha_s^2} \end{aligned}$$

(2). The relation for tangent (\bar{Y}_5/X_d^2) of MFL in side of f_s .

$$\frac{\bar{Y}_5}{X_d^2} = \frac{Y_5}{X_d^2} \cdot \frac{(X_p/X_d)^2}{(X_p/X_d)^2 - Y_5/X_d^2} \quad (A-2)$$

$$\text{where } \bar{Y}_5 = Y_5 \left[X_p^2 / (X_p^2 - Y_5) \right]$$

(3). The relation for tangent ($\bar{u}X_d^2$) of MFL in side of f_p .

$$\frac{1}{\bar{u}X_d^2} = \frac{1 - uX_d^2(X_p^2/X_d^2)}{uX_d^2} \quad (A-3)$$

$$\text{where } \bar{u} = u / (1 - uX_p^2), \quad u = 1/Y_5$$

(4). The relation for decision of R_5 .

$$\begin{aligned} \frac{R_5}{X_d} &= -\frac{1}{2} \left(\frac{M}{X_d} - \frac{N}{X_d} \right) - \frac{R_d}{X_d} \\ &\quad + \sqrt{\left[-\frac{1}{2} \left(\frac{M}{X_d} - \frac{N}{X_d} \right) \right]^2 + \frac{N}{X_d} \frac{M}{X_d} - \left(\frac{X_p}{X_d} \right)^2} \quad (A-4) \end{aligned}$$

$$\text{where } \frac{M}{X_d} = \left| \frac{V_0(1)}{V_1} \right|_{\max}, \quad \frac{N}{X_d} = \left| \frac{V_0(1)}{V_1} \right|_{\min}.$$

(5). The relation for X_s .

$$\frac{X_s}{X_d} = \frac{1}{2} \cdot f_s \cdot \frac{X_p/X_d}{\Lambda} \quad (A-5)$$

$$\text{where } \Lambda = f_p - f_s$$

(6). The relation for R_s .

$$\frac{R_s}{X_d} = \frac{(X_p/X_d)^2 [M/X_d + (R_5 + R_d)/X_d]}{(M/X_d)^2 - (R_5 + R_d)^2/X_d^2 - (X_p/X_d)^2} \quad (A-6)$$

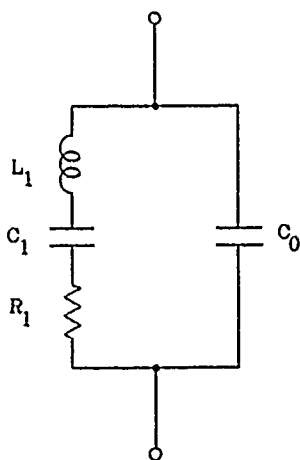


Fig.A-1 Equivalent circuit of quartz crystal unit.

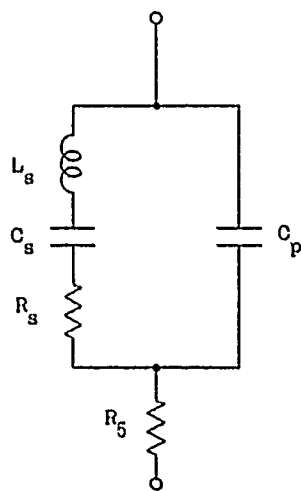


Fig.A-2 New equivalent circuit of quartz crystal unit.

REPORT ON THE WORKSHOP ON
TRACEABILITY OF QUARTZ MEASUREMENTS
TO U.S. STANDARDS

John A. Kusters

Ball Corporation, Efratom Division
Irvine, California 92715

Summary

A workshop on the traceability of quartz resonator measurements to the U.S. National Standards using the S-parameter measurement methods as outlined in ANSI/EIA-512-1985¹, was held at the Boulder facility of the National Bureau of Standards on January 15-16, 1986.

The purpose of the workshop was to define a methodology whereby measurements of quartz resonators could ultimately be made traceable to U.S. and other national measurement standards.

Because of limited resources, NBS does not currently have the capacity to calibrate complete user crystal measurement systems. After lengthy discussion, a plan proposed by NBS was adopted in principle by the workshop attendees. This plan calls for the use of commercially available 7 mm. precision shorts, opens, load terminations, and thru-lines whose impedances are determinable from first principles and/or have been calibrated using methods traceable to the national standards.

NBS, working with industry representatives, would undertake to develop transitional coaxial sections to permit use of standard 7 mm. calibration impedances on normal crystal mounting adapters.

Measurement Method

ANSI/EIA-512-1985 defines standard methods for the measurement and determination of the equivalent electrical parameters of quartz crystal resonators over the frequency range from 1 kHz. to 1 GHz. The standard is based on determining the motional-arm parameters of a crystal resonator from an admittance characteristic obtained by curve fitting a circle to a series of data points calculated from S-parameters measured on the unit as a function of frequency.

The standard is concerned only with the equivalent electrical representation of a quartz resonator in a narrow range of frequencies in the vicinity of the desired mode of vibration at a given excitation level. Non-linear amplitude behavior or the presence of unwanted vibrational modes makes the measurement invalid.

Two basic methods are described; the first uses a two-port transmission method which characterizes the crystal unit as a three-terminal network. The second is a one-port reflection method which characterizes the unit as a two-terminal device. Both yield the same motional parameters (R_1 , C_1 , and L_1) but give different representations for the static capacitance parameters. The simplified equivalent circuit representations are shown in figure 1.

Both methods imply the existence of a suitable network analyzer and companion S-parameter test set,

such as the HP-3577A and HP-35675A. Many other commercially available test systems are also available.

Transformation from S-parameters to equivalent crystal parameters is relatively straightforward and fully outlined in ANSI/EIA-512-1985. The problem lies in calibrating the measurement system and determining the necessary correction factors to the measured data. For example, to fully implement the methodology of ANSI/EIA-512-1985 requires the determination of three system "error" terms for one-port measurements, and twelve "error" terms for two-port measurements. Each error term is complex and also a function of frequency.

Calibration

Although commercially available software² is available to handle the system of complex equations and manage the calibration and data reduction for both one- and two-port measurements, the major problem lies in calibration standards that are traceable to the national standards, or are determinable from first principles. Precision shorts, opens, load terminations, and thru-lines are also commercially available in many standard configurations from a variety of vendors. None of the standard configurations match those of typical crystal mounts.

Use of the methodology described in the standard requires a measurement head which will accept the desired crystal mount configuration. In virtually all circumstances, the crystal pin layout is such that only one pin can be accessed at a time using existing calibration standards. In addition, a crystal pin is rarely adaptable to a standard connector configuration. To attempt to solve these problems, many calibration adapters have been devised which mate properly with the crystal mount socket, and which in principle can be calibrated and traceable to existing national standards. However, many different crystal mounts are in use today. Any attempt to define a proper set of calibration standards for each of these would be a long and laborious task. Further, calibration traceable to national standards of each of these would exceed the resources of any national standards laboratory.

Workshop Discussion

On January 15-16, 1986, a workshop to discuss the traceability of crystal measurements to U.S. national standards was held at the Boulder facility of the National Bureau of Standards. Attendees included representatives from the National Bureau, the U.S. Army, the EIA P-11 Technical Working Committee, and other quartz industry members.

Included in the workshop were demonstrations of crystal parameter measurements using commercially available network analyzers and software, and custom crystal test mounts and calibration standards developed by members of the workshop.

Two key items were considered. The first concerned available calibration resources which would provide traceability of quartz resonators to existing national standards. This arises because of the need to have good measurement correlation between one measurement site and another, or between different companies. The second addressed a question of how precise the S-parameters need be known in order to achieve a desired level of accuracy and precision in the final parametric values.

Calibration

The workshop discussion centered on how the extremely wide variety of possible calibration adapters, standards, etc., that could be constructed to measure crystals, could be made in a manner that would allow traceability to the U.S. standards. The general conclusion was that this was an almost overwhelming task.

The workshop then examined a variety of precision shorts, opens, terminations, and thru-lines which are readily available commercially and are routinely calibrated using methods which provide the necessary traceability. The general consensus was that any method devised must be based on these calibration components and not on custom-designed adapters and standards.

Upon recommendation from the National Bureau, the workshop agreed to concentrate on calibration standards which are based on the precision 7 mm. (APC-7) connector. Because of the pin spacing on the crystal mount, any test head constructed must be designed to permit splitting the head into two, with sufficient space between test pins to accommodate the APC-7 calibration standards.

NBS, working with quartz industry representatives, will specify the design of a transitional section which will permit coupling of the APC-7 style calibration standards to the test head socket. This transitional adapter could be calibrated by NBS through its usual channels for calibration services.

Accuracy

Of interest also in the measurements of the quartz crystal is the accuracy required in the measurement system to achieve a desired level of confidence in the resonator parametric values. Given the system of complex equations which are a function of frequency and which must be solved to determine the crystal parameters, how good must the network analyzer and associated switches, test sets, test heads, cables, etc. be to get the desired results. A corollary might be: do we really need all those system "error" correction terms?

This becomes a question which can be approached either experimentally or analytically. Correlation of data obtained from a variety of measurements made on different systems with a variety of crystal cuts, overtones, etc., gives a good measure of repeatability, but no indication of accuracy. THE

standard quartz crystal doesn't exist, and if defined, probably couldn't be duplicated anyway. Further, the methods outlined in ANSI/EIA-512-1985 present the best available today. In other words, no calibration standard exists against which the experimental results could be judged.

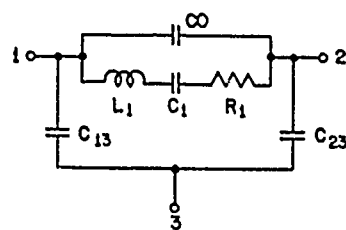
Because of the above, several universities were approached to undertake as a graduate thesis project the analysis of the system of equations, correction terms, etc., using an analytical approach. By generating the necessary computer models and exercising them over a variety of conditions, the sensitivity of the parametric values of the resonator to uncertainties in measurement calibration should be readily determined. The EIA Technical Working Committee is currently evaluating a proposal to undertake this task.

Conclusion

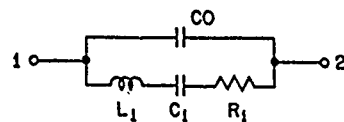
This was a preliminary report which outlined current status of the work being done between NBS, the EIA P-11 Technical Committee, and the quartz industry to provide a methodology for the measurement of quartz crystals that is consistent, repeatable, and traceable to U.S. and other national standards.

References

1. Electronic Industries Association, Standard Methods for Measurement of the Equivalent Electrical Parameters of Quartz Crystal Units, 1 kHz to 1 GHz, ANSI/EIA-512-1985, April 1985, Washington DC, 20006.
2. For example, Solution #9, Innovative Measurement Solutions, Atlanta GA, 30307.



(a) TWO-PORT REPRESENTATION



(b) SINGLE-PORT REPRESENTATION

Figure 1. Equivalent electrical networks commonly used to represent a piezoelectric crystal unit over a narrow frequency range near an isolated mode of vibration.

A FREQUENCY ADJUSTABLE ULTRA-COMPACT, HIGH-PERFORMANCE QUARTZ CRYSTAL OSCILLATOR
AND ITS SIMPLE TEMPERATURE COMPENSATION METHOD

O. OCHIAI, F. TANURA and Y. MASHIMO

SEIKO ELECTRONIC COMPONENTS LTD.
1110 Hirai-cho, Tochigi-shi, Tochigi-ken, Japan 328

Summary

A very small-size, high-precision quartz crystal oscillator has been developed. Perhaps its most obvious application is as a reference digital source in communications equipment, e.g., mobile radio systems requiring very high accuracy. This oscillator's frequency is adjustable electrically by external means. The oscillator comprises of a highly-miniaturized GT-cut quartz crystal resonator, a floating electrode MOS variable capacitor, and a CMOS-IC. This new semiconductor element, dubbed the PEAC, has been commercially introduced as a high accuracy oscillation frequency adjustor and capable of maintaining a constant capacitance over long time. Namely, it is capable of varying oscillation frequency continuously within 7 to 25 pF by applying voltage (+14V to +16V). Employment of the PEAC into this oscillator allows for complete encapsulation of the device and for high reliability. Frequency stability is less than 2.5 ppm (from -30°C to +70°C); and dimensions are 12.5mm x 12.5mm x 6.5mm.

Introduction

Quartz crystal oscillators integrated with an oscillating circuit are indispensable to various types of high-tech electronic equipment. Recently, their application has begun to diversify from communications equipment, television sets and clocks, to video cameras, video discs, compact discs and optical communication instruments. Particularly, the strong demand for the oscillators featuring an ultra-compact, simple-structure and also very low current consumption is now increasing.

In order to best respond to the above requirements, we have developed a 2-MHz very small-size, high-precision quartz crystal oscillator. It incorporates a miniaturized GT-cut quartz crystal resonator with superb frequency temperature characteristics, a newly-developed floating electrode MOS variable capacitor that is capable of adjusting capacitance, and CMOS IC. Also, a new-temperature compensation method using this variable capacitor element was worked out.

Structure of a Floating Electrode MOS Variable Capacitor

A new semiconductor element, dubbed the PEAC, has been developed. It is capable of varying capacitance electrically and also maintaining a constant capacitance. In the conventional variable capacitors, frequency adjustment has been carried out by rotating electrode mechanically. So, there are the following drawbacks:

1. Lacking in long-term frequency stability; and
2. Making holes in the case is needed when adjusting the frequency by external means, and thus complete encapsulation is not possible.

The PEAC is just a wholly-new semiconductor element that was realized as a result of surmounting the above drawbacks.

Figure 1 outlines its structure. The PEAC comprises of three terminals--namely, variable injection terminal (T_1), capacitance terminal (T_C), and ground terminal (T_G). The capacitance between T_C and T_G is used as a variable capacitor. In order to vary and determine the capacitance value arbitrarily, the floating electrode covered with silicon dioxide thin film, which is just like a non-volatile memory IC, is used.

When applying the high positive voltage (+14 to +16V) between T_1 and T_G , electric charge will be removed through oxide thin film from the floating electrode. This results in increase in capacitance between T_C and T_G . On the contrary, when applying negative voltage (-14V to -16V) between T_1 and T_G , electric charge will be injected through the oxide thin film into the floating electrode. This results in decrease in capacitance between T_C and T_G ; as far as high voltage is not applied to T_1 terminal, electric charge injected into the floating electrode can be retained in long-term and also the capacitance between T_C and T_G can be kept constant. Accordingly, once the capacitance is adjusted, it is not necessary to apply voltage to T_1 terminal.

Oscillation frequency variation characteristics

Figure 2 shows the oscillating circuit. To adjust oscillation frequency, the PEAC is inserted into Gate side of the inverter. Ground terminal T_1 is connected to the positive power supply (+ 5 V in usual), and capacitance terminal T_2 is connected to Gate of the inverter. Floating electrode terminal T_3 , which is out of the case of the oscillator serves as the adjustment terminal of oscillation frequency. Symbols used are as shown in the figure.

Figure 3 shows oscillation frequency variation characteristics. Oscillation frequency will vary when positive voltage is applied between the positive power supply (5V in usual) and the frequency adjustment terminal. Applied voltage is 100 msec pulse. The abscissa denotes the number of pulses, and the ordinate denotes frequency deviation, where 2.1 MHz is Oppm. When negative voltage is applied to the adjustment terminal, the internal capacitance will increase, whereas the oscillation frequency will decrease. The higher the voltage is applied, the faster the rate of change in the frequency becomes, because frequency variation depends on the applied voltage and the time period.

Figure 4 also shows oscillation frequency variation characteristics. Frequency can easily be varied by applying voltage via external means and also varied in a wide range of 80 ppm and more. Practically, frequency can be adjusted by applying optimum voltage for several tens ms, depending on deviation from the desired frequency.

A 2.1-MHz miniaturized GT-cut resonator

A newly developed 2.1 MHz miniaturized GT-cut resonator, the MGQ, is used in this oscillator. Figure 5 shows its shape. It is well known that outstanding frequency temperature characteristics over wide temperatures by coupling a Width Extensional Mode and a Length Extensional Mode. As shown in figure, the vibration part and supporting part of the MGQ are integrated with our unique photolithographic technology.

The strongest advantage of the resonator is excellent frequency temperature characteristics of +5 ppm or less from -30°C to +70°C. However, ratio of the length and width must be strictly controlled to attain such temperature stability. Therefore, a new temperature compensation method using the PEAC was worked out to realize the high-precision oscillator.

Study of temperature compensation method using PEAC

The capacitance can be varied by applying high voltage to the adjustment terminal and also kept constant over long time unless adding the voltage further. Oscillation frequency variations are as shown in the figures, when varying direct bias voltage between T_1 and T_2 .

The measurement circuit is shown in Figure 6. Bias voltage V_b is applied between ground terminal T_1 and capacitance terminal T_2 via resistor R as a reference of power supply voltage.

Oscillation frequency variations versus bias voltage V_b are shown in Figure 7. The ordinate denotes oscillation frequency deviation, when setting to Oppm at 0.5V bias voltage. Frequency varies almost linearly at about 0.3 V to 1.2 V of bias voltage. This means that the PEAC can be used for temperature compensation as the same as a conventional variable capacitor diode. In this connection, we studied a simple temperature compensation method using the PEAC and a thermistor.

Figure 8 shows the circuit diagram of the temperature-compensated oscillator which was trially manufactured. Output voltage of the oscillator built-in voltage regulator is about 1.8 V as a reference of + 5 V_{DC} positive power supply. Bias voltage V_b to be applied to the PEAC is determined with a ratio of composite resistor Rx including both thermistor and solid resistor RI.

In this circuit, as temperatures rise, bias voltage of the PEAC will increase, and oscillation frequency will become high. Therefore, temperature characteristics of the resonator with a primary negative temperature coefficient can be compensated.

Experimental results

Figure 9 shows experimental results of Figure 8. Full line 1 shows temperature characteristics before temperature compensation. Broken line 2 is temperature characteristics obtained by calculating characteristics of the PEAC and values of a composite resistor Rx including a thermistor. Full line 3 represents temperature characteristics after temperature compensation. Although a very simple temperature compensation circuit is employed, temperature characteristics after compensation are excellent--namely, less than + 0.5ppm are obtained from -30°C to +80°C.

This example is only one case when temperature characteristics with a primary negative temperature coefficient is compensated. In the same way, temperature characteristics with a primary positive temperature coefficient are easy-to-compensate. Furthermore, second-and third temperature coefficients of the GT-cut quartz crystal resonator are very small compared to these of an AT-cut quartz crystal resonator and it allows for easy temperature compensation.

Specifications for the temperature-compensated quartz crystal oscillator

The following are specifications for this quartz crystal oscillator:

- | | |
|---------------------------------|---|
| 1. Frequency f_0 : | 1.7 MHz to 3.0 MHz
or divided output |
| 2. Frequency stability: | less than + 2.5 ppm |
| 3. Operating temperature range: | from - 30°C to +75°C |
| 4. Frequency warp range: | more than 10 ppm |
| 5. Current consumption: | less than 1 mA |
| 6. Case size: | as shown in Figure 10 |

Conclusion

Experimental data we introduced here shows that the floating electrode variable capacitor element, the PEAC, is capable of adjusting frequency easily and accurately by applying voltage and also of compensating temperature by combining simple circuits including a thermistor.

Employment of both this floating electrode capacitor element and the miniaturized GT-cut quartz crystal resonator has realized a 2-MHz wholly-new, very small-size, high-precision quartz crystal oscillator.

This oscillator can respond to today's increasingly sophisticated needs as a reference signal source in diverse fields, such as mobile radio communication equipment.

Acknowledgment

The authors would like to thank Kubota, and Morizono for their help and encouragement and also Hattori of Seiko Instruments & Electronics Ltd. for appropriate technical advices, and suggestions.

References

1. Y. Hattori, R. Matsuzaki, K. Tanaka and A. Watanabe: "Application of Floating MOS Variable Capacitor for Watch ICs," CIC, pp 9 to 12, 1984
2. K. Kawashima, H. Sato, and O. Ochiai: "New Frequency Temperature Characteristics of Miniaturized GT-Cut Quartz Resonators," Proc. 34th AFCS, pp 131 to 139, 1980
3. O. Ochiai, A. Kudo, A. Nakajima, and H. Kawashima: "A Method of Adjusting Resonant Frequency and Frequency Temperature Coefficients of Miniaturized GT-Cut Quartz Resonators," Proc. 36th AFCS, pp 90 to 96, 1982.

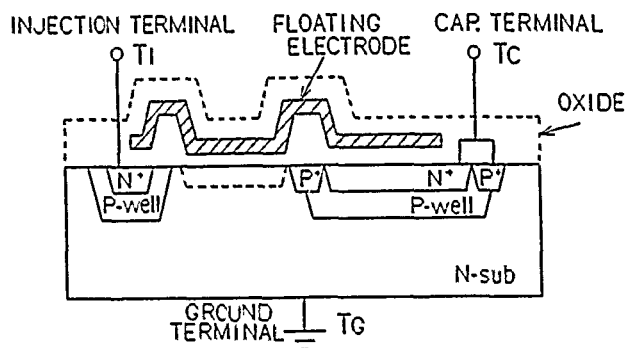


Fig. 1 Structure of the PEAC

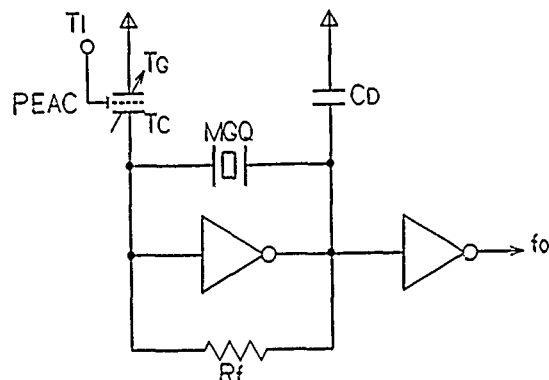


Fig. 2 Oscillator Circuit

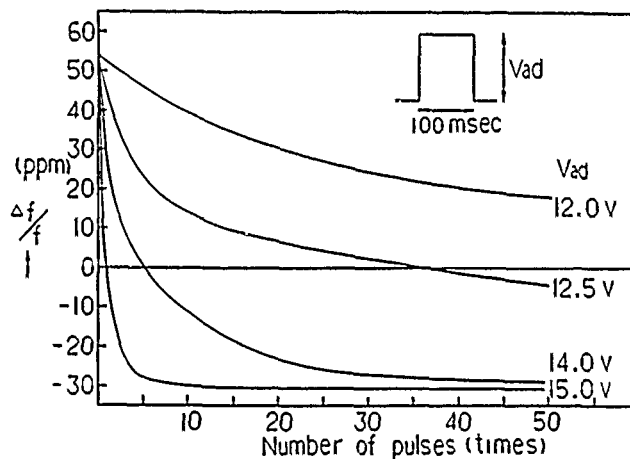


Fig. 3 Frequency Variation Characteristics

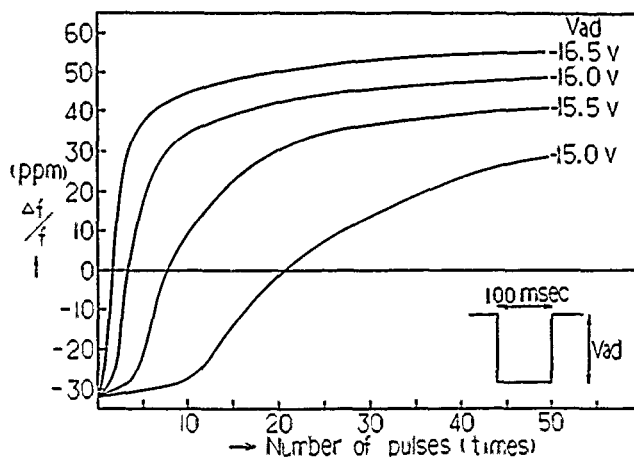


Fig. 4 Frequency Variation Characteristics

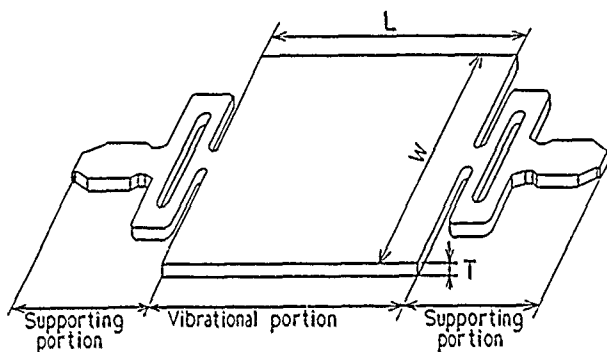


Fig. 5 Shape of miniaturized GT-cut quartz resonator

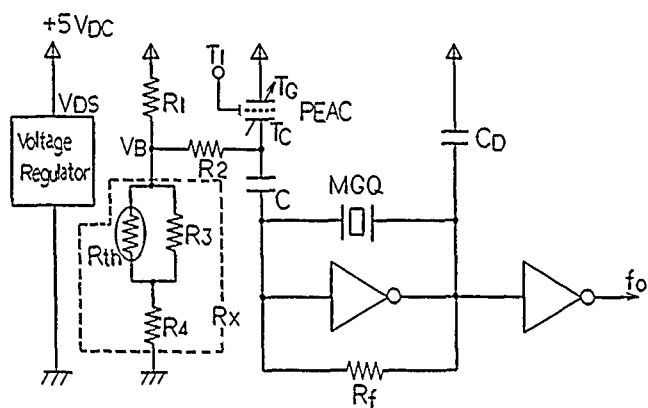


Fig. 8 Temperature Compensated Oscillator Circuit

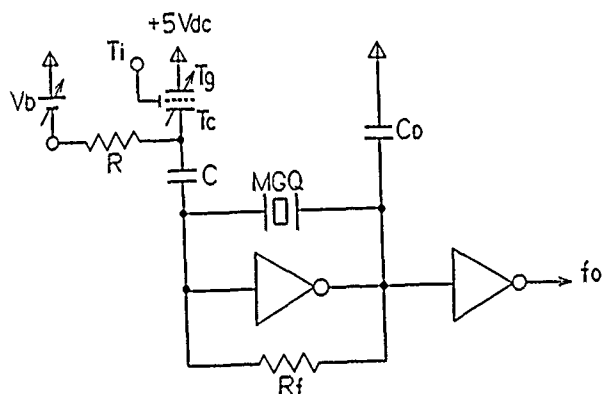


Fig. 6 Measurement Circuit

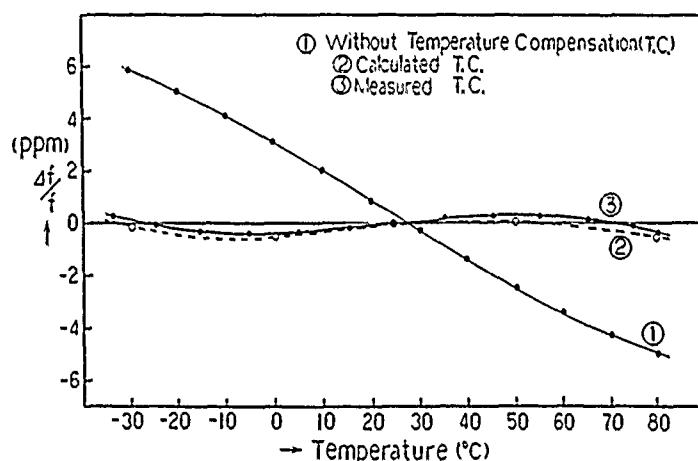


Fig. 9 Frequency Temperature Characteristics

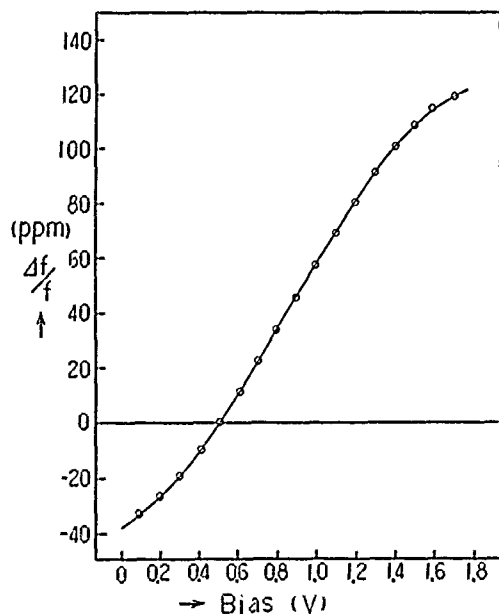


Fig. 7 Frequency vs. Bias Voltage of the PEAC

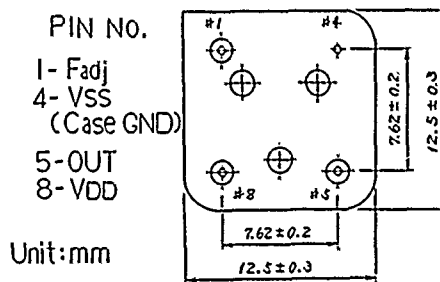
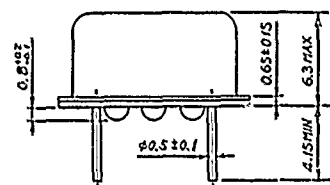


Fig. 10 Case dimensions

Low Noise Crystal Oscillators Using 50-ohm, Modular Amplifier Sustaining Stages

M.M. Driscoll
Westinghouse Defense and Electronics Center
Baltimore, Maryland 21203

SUMMARY

This paper describes several crystal oscillator configurations that use self-contained, 50-ohm, modular amplifiers as the oscillator sustaining stage. Broadband modular amplifiers, often packaged in TO-8 enclosures, are available that exhibit adequately high (20 dB) gain, high reverse isolation, and extremely low flicker-of-phase noise, and makes them attractive candidates for use in low noise crystal oscillator circuitry.

Two oscillator configurations have been evaluated using VHF, overtone, AT and SC-cut quartz crystal resonators: (A) with the resonator operated at series resonance, and series connected as part of a resistive pad in the oscillator positive feedback circuit path, and (B) with a shunt resonator, operated at series resonance and connected across the amplifier input terminals via a quarter wavelength transmission line.

An advantage of these configurations is that accurate, independent measurement of 50-ohm sustaining stage amplifier phase noise at operating point drive levels is easily made. These measurements indicate reliable attainment of amplifier 1/f phase noise levels of -160 dB/Hz at 100 Hz carrier offset. In the VHF oscillators, the contribution of amplifier phase noise for fm < 1 KHz is more than 10 dB below which result from typically obtained crystal resonator short-term frequency instability. At higher modulation rates the output signal noise floor is determined by the amplifier effective noise figure and input drive level. Noise floor levels of -173 dB/Hz have been obtained for type A configuration oscillators employing SC-cut resonators operating with 6-8 mw crystal dissipation and 3dBm amplifier input drive. For the type B configuration, a noise floor level of -170 dB/Hz was obtained for 2.5 dBm (AT-cut) crystal dissipation and -3 dBm amplifier input drive. The low effective noise figure of the type B sustaining stage amplifier appears to be associated with the extremely low value of out-of-band impedance at the amplifier input resulting from the quarter wavelength line.

The multiple decade bandwidth available using hybrid and MMIC, 50 ohm amplifiers indicates these circuits will operate quite satisfactorily with UHF bulk wave and SAW resonators.

INTRODUCTION

Figure 1 shows, in block diagram form, two convenient methods for both analyzing and

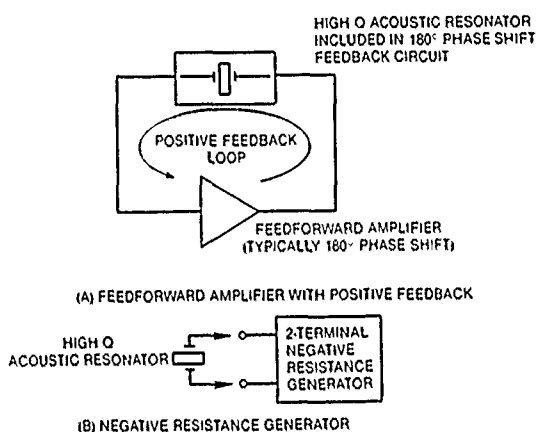


Figure 1. Basic Oscillator Configurations

designing crystal-controlled harmonic oscillator circuits. As shown in the figure, the oscillator sustaining stage (active device) can be considered as: (a) a feed forward amplifier with the oscillator resonator or frequency-determining element included in a positive feedback circuit, or (b) a two-terminal, negative resistance generator to which the resonator is connected. For the purpose of this paper, it is convenient to view the circuit as a feedforward amplifier, as shown in figure 1(a).

In the circuit of figure 1(a), the steady-state conditions for oscillation are unity closed loop gain and 2π radians closed loop phase shift at the operating frequency. To initiate oscillation, the circuit must exhibit a small signal excess gain (usually on the order of 2 to 6 dB). The large signal, or steady-state unity gain requirement is met by employing auxiliary AGC or ALC circuitry, or by allowing the signal level to increase until amplifier nonlinearity results in closed loop gain decrease to unity. Often the amplifier is biased to current limit at an established quiescent signal level.

Figure 2 shows how the sustaining circuit (open loop) phase noise sideband spectrum is related to that of the (closed loop) oscillator output signal. At frequencies less than the half-bandwidth of the resonator, there is a conversion of sustaining stage open loop flicker-of-phase (10 dB/decade) noise to flicker-of-frequency (30 dB/decade) noise in the oscillator output signal. This conversion of open loop signal phase perturbation to closed loop frequency perturbation is a consequence of the circuit maintenance

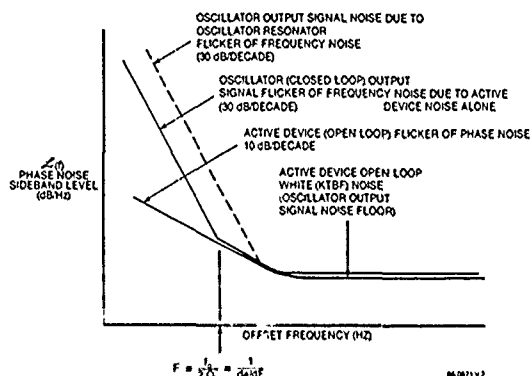


Figure 2. Typical Oscillator Output Signal
Phase Noise Sideband Spectral Character

of constant ($2\pi\Omega$) close loop phase shift, and is related to the closed loop signal group delay (resonator phase slope) which is proportional to resonator loaded Q^1 . For a single resonator:

$$\frac{d\phi}{df} = \frac{2Q_L}{f_0} \quad (\text{radians/Hz}) \quad [1]$$

The solid curves of figure 2 are valid so long as the short-term stability of the passive resonator itself can be neglected. Recent measurements conducted at Westinghouse show that, especially with the use of sustaining stage circuitry exhibiting very low flicker of phase noise, resonator flicker of frequency noise is dominant using VHF quartz crystal (bulkwave) resonators, as shown by the dashed curve in figure 2². Other investigators have shown this situation also can exist with the use of HF quartz crystals as well as UHF surface acoustic wave resonators.³⁻⁶

Clearly the following design criteria need to be met to the greatest degree possible in connection with the fabrication of highly reliable, easily producible, low noise RF oscillators.

- (1) Use of sustaining stage circuitry exhibiting very low flicker-of-phase noise and low noise figure.

- (2) Use of crystal resonators exhibiting best possible short-term frequency stability.
- (3) Minimization of resonator loading. (However, when the effect of sustaining stage 1/f noise on output signal near carrier noise is negligible compared to actual resonator short-term instability, higher degree of loading is tolerable.)
- (4) Use of highest possible RF drive (consistent with items 1 and 2 above) in order to obtain maximum S/N ratio in connection with active device white (KTBF) noise spectrum which dominates at high modulation rates (i.e., noise floor level).
- (5) Use of circuit components (especially the resonator and the sustaining stage

amplifier) whose short-term frequency stability or phase noise characteristics can be easily and accurately evaluated before use in the oscillator circuit. In this respect, use of a single self-contained sustaining stage amplifier is desirable.

MODULAR AMPLIFIER FLICKER-OF-PHASE NOISE

Figures 3 and 4 show a representative comparison of the flicker-of-phase characteristics measured for two different self contained, modular, multiple decade bandwidth RF amplifiers that are representative of the many types of units in widespread usage in the industry.

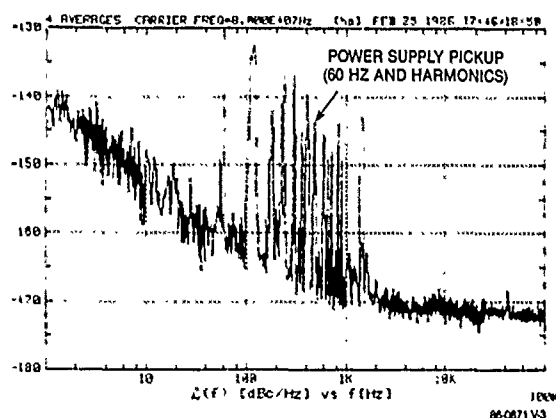


Figure 3. Measure Phase Noise Sideband Levels for Q-Bit, QBH-125 Amplifier at 1 dB and 4 dB Gain Compression (Spectra Are Near-Identical)

As can be seen from figure 3, some amplifier types exhibit very low 1/f phase noise when driven at levels up to an including 3 dB gain compression. Figure 4 shows, by comparison, that amplifier operation well into gain compression can result in substantially increased flicker of phase noise (and noise floor degradation) for some alternative

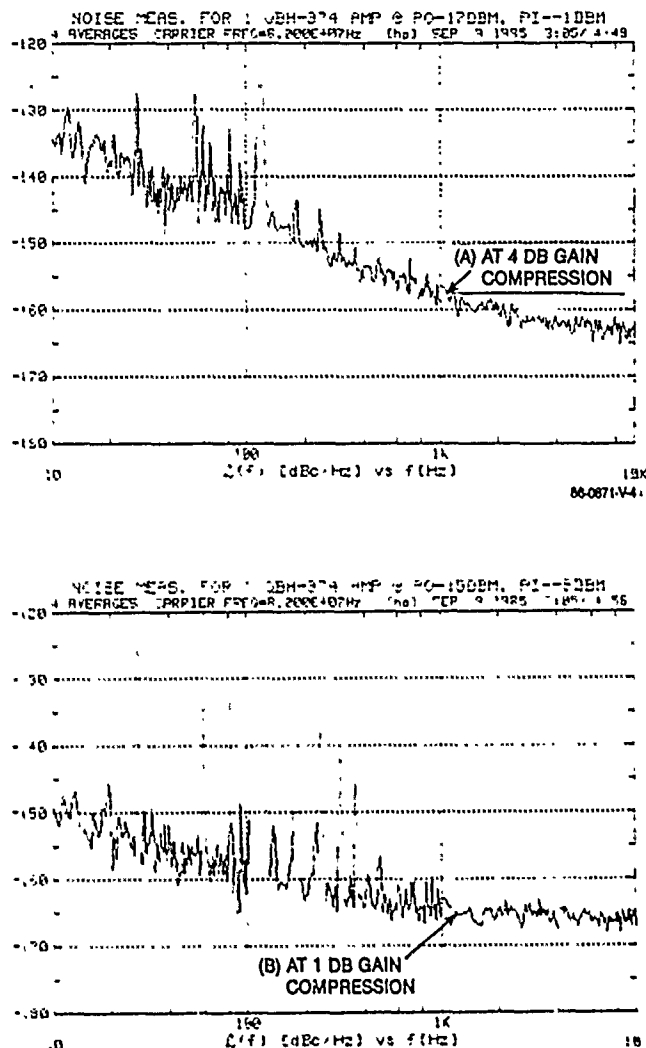


Figure 4. Phase Noise Spectrum for QBH - 374 Modular Amplifier

amplifier types. Unfortunately, this aspect of device performance is not specified by amplifier manufacturers.

In general it appears that amplifiers incorporating emitter degeneration (external RF emitter resistance) tend to exhibit adequately low $1/f$ phase noise. This is consistent with observations first made by D. Halford in 1968.⁷

Self contained, 50 ohm, monolithic amplifiers recently have been developed which exhibit GHz operating bandwidth, low propagation delay, and supply voltage-variable compression point.⁸ Although per-stage gain currently is limited to 10 to 12 dB, measurements conducted at Westinghouse indicate that these devices also inherently exhibit low flicker-of-phase noise (likely associated with the use of external RF emitter resistance). Use of this class of amplifier (possibly including a cascade of 2 devices) might be considered for use in oscillators incorporating high VHF and UHF resonators.

The implication of the figures 3 and 4 data is that if these devices are used as a self-limiting oscillator sustaining stage, the oscillator small signal excess gain (equal to the degree of gain compression in the sustaining stage amplifier for steady state operation) must be held to 2 dB or less. Although this can be accomplished easily in a laboratory environment, circuit operation over military temperature extremes demands use of increased small signal excess gain to ensure successful circuit oscillation. Based on this fact, Schottky diodes were used in the oscillators to provide RF signal limiting and allow amplifier use just below gain compression.

Figure 5 shows a schematic diagram of oscillator circuit that was designed to operate using a third overtone, 80 MHz, SC-cut crystal resonator. As shown in the figure, the circuit consists of a 20 dB gain modular amplifier, a

tuned circuit, a power divider, a Schottky diode (limiter), and the crystal resonator which, operated at series resonance, forms part of a resistive pad in the oscillator positive feedback circuit path.

The Schottky diodes provide RF signal limiting and allow amplifier operation below the compression point. Resistor R1 value is selected to present a matched (50 ohm) impedance to the power divider. Resistors R2, R3 are selected in conjunction with the quartz crystal series resistance, to provide desired small signal excess gain (4 to 5 dB), crystal dissipation (7 dBm), and amplifier input drive (2 dBm).

The tuned circuit must exhibit enough selectivity to suppress oscillation at the SC-cut crystal B mode resonance approximately 10% above the operating frequency. A tuned circuit bandwidth of 10% to 15% suffices. Ideally, the tuned circuit also should exhibit a signal phase shift at the point of minimum loss equal to and opposite of that of the amplifier. Because of amplifier delay, the phase shift is not 180 degrees. For the circuit of figure 5, the amplifier phase shift is approximately 150 degrees. An equal and opposite tuned circuit phase shift can be obtained using the circuit of figure 5 by selecting $\arctan(\omega L5/RL) \approx \arctan(\omega L3/RS) \approx 150/2 = 75$ degrees, where RS and RL are the tuned circuit source and load impedances (i.e., 50 ohms). R4 is included in the circuit to provide a matched tuned circuit output impedance of 50 ohms (at the operating frequency) so that maximum power divider input output isolation is achieved. Use of R4 is required as a result of losses in L3, L4 and L5. The tuned circuit additionally affords excellent harmonic rejection. Output signal harmonics for the circuit of figure 5 were measured below -60 dBc. Also shown in figure 5 is a simple, alternative method for providing adjustable, Schottky diode limit level.

Figure 6 shows the phase noise sideband spectra measured for two prototype oscillators

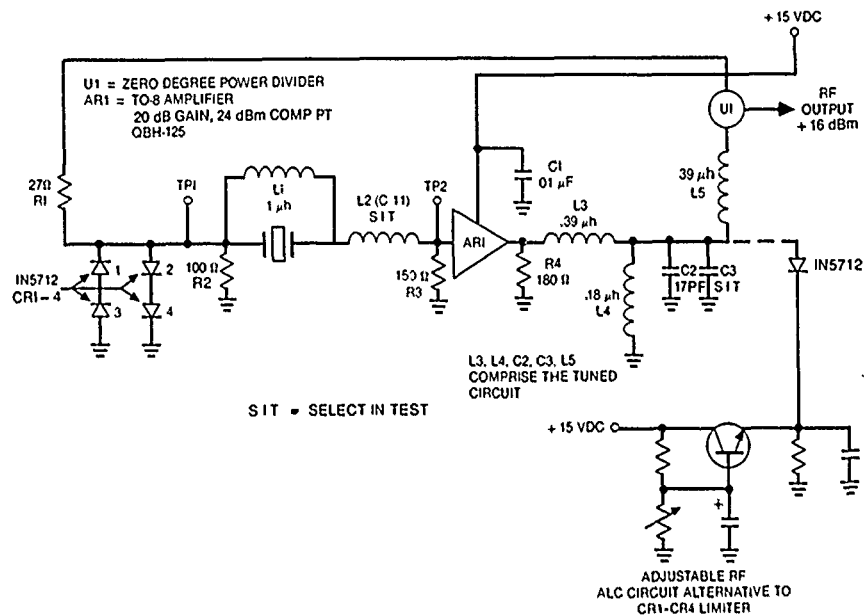


Figure 5. 80 MHz, SC Cut - Crystal Controlled Oscillator (Type A Circuit)

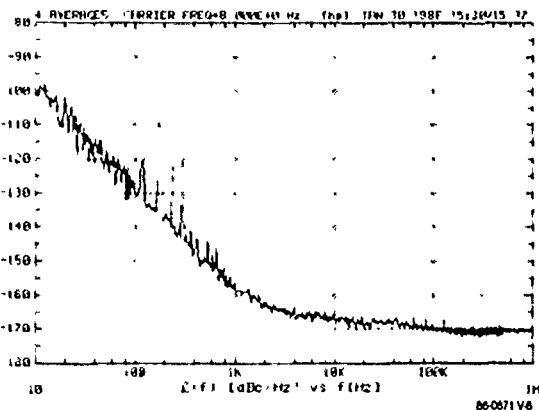


Figure 6. Type A Oscillator Performance Using An SC-Cut Crystal

operating at 80 MHz. Figure 7 shows the results of independent measurement of crystal short term stability. The figure 7 measurements were made by driving each crystal, in a phase bridge, from a common signal generator into a phase detector.^{1,2} In figure 7, the measured phase noise sideband level of -145 dB/Hz at $f_m = 100$ Hz corresponds to a phase noise spectral density of 6.3×10^{-5} rad²/Hz. In the phase bridge the transmission phase slope was 2×10^{-3} rad/Hz. Therefore, one can convert the

measured phase noise to a crystal resonator frequency noise of 1.6×10^{-9} Hz²/Hz which corresponds to a predicted combined noise sideband level ($f_m = 100$ Hz) of -131 dB/Hz for two oscillators incorporating each of the two crystals. As seen in figure 6 this is within 1 dB of that measured for the two prototype

oscillators. The oscillator near carrier noise results from crystal instability and not amplifier phase noise. The figure 6 noise floor level of -170 dB/Hz for two oscillators (-173 dB/Hz for each oscillator) is consistent with 2dum amplifier drive and 3 dB noise figure. The circuit contains no adjustable elements and approximately 1/3 to 1/2 the number of components employed in previously used production crystal oscillator circuits at Westinghouse. The use of limiter diodes result in reliable circuit operation (more excess gain) over a wider temperature range.

The circuit of figure 5 also has been used with 5th overtone AT-cut crystals, operating at reduced levels of crystal dissipation. This was accomplished by shorting Schottky diodes CR3 and CR4 to reduce the quiescent RF operating signal level, and by substituting a lower compression point amplifier for AR1. The resulting oscillator signal phase noise sideband spectra, using an AT-cut resonator, is shown in figure 8.

ALTERNATIVE OSCILLATOR CONFIGURATION

The author considered the use of alternative circuit configurations that would incorporate

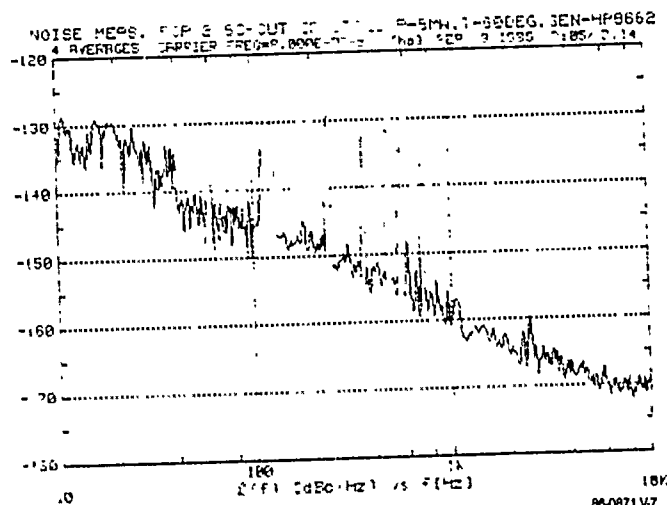


Figure 7. Measured Phase Noise Spectra for Prototype Oscillator SC-Cut Crystals at 5 mw Drive

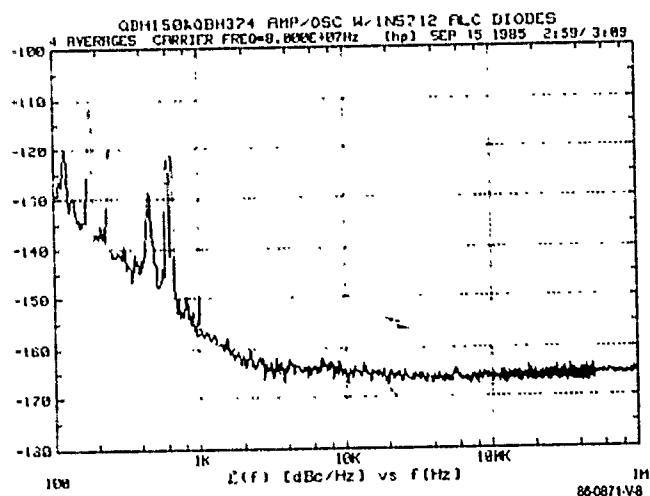


Figure 8. Measured Noise Spectrum for Two (Type A Circuit With CR3 CR4 Shorted) Oscillators Using AT-Cut Resonators With Amplifier Input Drive = -3 dBm

an impedance transformation circuit between the resonator and the 50 ohm amplifier impedances to provide additional design flexibility with regard to resonator loading (Q degradation), dissipation, and amplifier input drive level.

Use of RF transformers was not considered flexible enough owing to the fact that these devices are readily available only for specific impedance levels and turns ratio. It was decided that both the impedance inverting and impedance-transforming properties of a quarter-wave transmission line could be used to advantage. At high VHF and UHF frequencies, actual transmission line implementation in distributed form (Microstrip, etc) is practical. At HF and low VHF frequencies, the

line can be implemented in lumped element form. Further, the lumped element approximation allows design and use of transmission line characteristic impedance values not easily implementable in distributed form.

Figure 9 shows the configuration that was used. The circuit is similar to that of figure 5 except that the (AT-cut) crystal resonator, operated at series resonance, is connected across amplifier input terminals via a 5 pole, lumped element approximation, quarter wavelength transmission line having $Z_0 = 24$ ohms. The crystal, having 45 ohm series resistance, is connected in shunt at one end of the transmission line.

The transmission line essentially transforms the series resonance characteristic ($R_s = 45$ ohms) of the crystal to a parallel resonant characteristic across the amplifier input. The parallel resonant resistance presented by the line is $Z_0^2/R_s \approx 13$ ohms. Since the parallel combination of amplifier input impedance (50 ohms) and feedback circuit impedance (89 ohms) is 32 ohms, or 2.5 times that resulting from the (transformed) crystal resistance, crystal Q degradation is on the order of 30% ($Q_L = 0.7 Q_u$).

The oscillator of figure 9 was operated with 1.8 mw AT-cut crystal dissipation and -3 dBm amplifier input drive. Measurements of output signal phase noise sideband spectra for two such oscillators (figure 10) show a noise floor (-167 dBc/Hz for two oscillators) that is approximately 2 dB below that expected, based on published ($R_s = 50\Omega$) amplifier 3 dB noise figure. This appears to be associated with lower effective amplifier noise figure at high ($f_m > 10$ kHz) carrier offset frequencies resulting from the extremely low out of band impedance of the quarter wavelength line at the amplifier input. Figures 10 and 11 also show that near-carrier noise level (100 Hz) = -129 dB/Hz for two oscillators) is a result of short-term instability in the resonators themselves. At 100 Hz carrier offset, the phase noise sideband level that would result from the measured amplifier open loop phase noise level of -160 dB/Hz is -145 dB/Hz per oscillator, based on a loaded crystal Q of $0.7 Q_u = 7 \times 10^5$. This is 12 dB below the phase noise sideband level that is attributable to resonator short-term frequency instability.

CONCLUSIONS

It has been shown that low (flicker-of-phase) noise, 50 ohm, modular amplifiers can be used as the sustaining stage in low noise, crystal-controlled oscillator circuitry.

Two very straight forward oscillator configurations have been fabricated, and measurement of output signal phase noise sideband spectra indicate (as predicted) the influence of amplifier flicker-of-phase noise is more than 10 dB below that of actual resonator short-term frequency instability. As such, the oscillator configurations provide excellent vehicles for resonator stability evaluation.

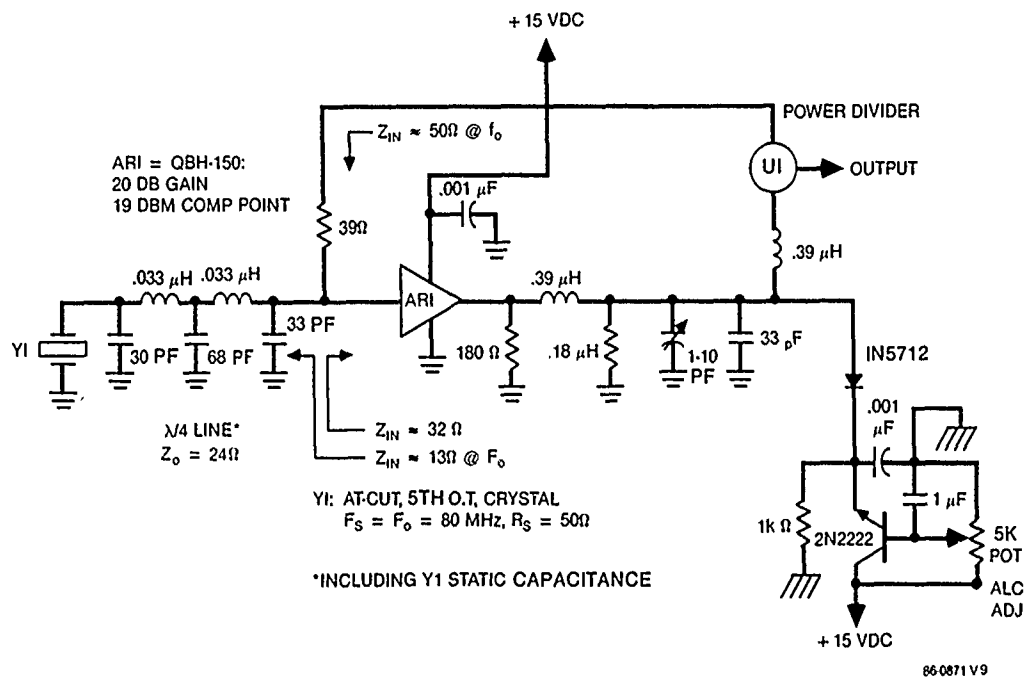


Figure 9. Alternative/Type B) Configuration Oscillator Using Lumped Element, Quarter Wavelengths Transmission Line

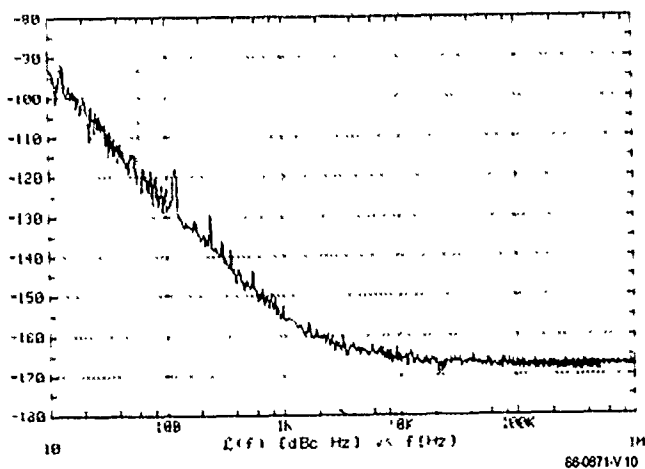


Figure 10. Measured Phase Noise Sideband Spectra for Two Type B Oscillators With -3 dBm Amplifier Input Drive

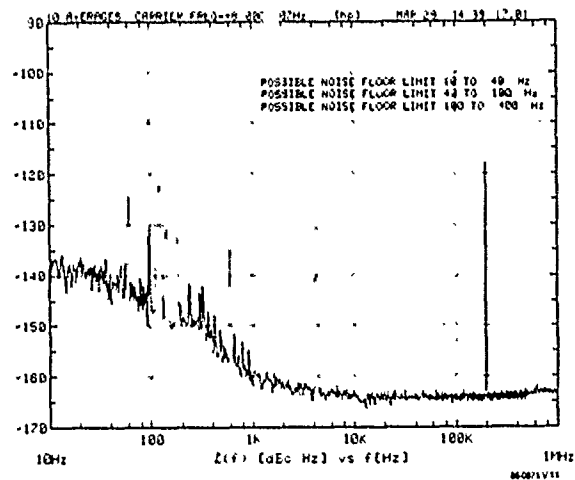


Figure 11. Measured Phase Noise Spectrum for Prototype Oscillator At-Cut Crystals at 2 mW Drive (Compare to Figure 7)

An advantage associated with the use of self-contained, 50 ohm amplifier, oscillator sustaining stage is that independent evaluation of amplifier phase noise performance is accomplished easily.

The basic design concept lends itself to use with other types of frequency control elements (such as SAW resonators) and with self-contained, MMIC, 50-ohm amplifiers.

REFERENCES

1. D.B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum", Proc. IEEE, Vol. 54, No. 2, Feb., 1966, pp 329-330.
2. M.M. Driscoll et al, "Spectral Degradation in VHF Crystal Controlled Oscillators Due to Short-Term Frequency Instability in the Quartz Resonator", Proc. 1985 IEEE Ultrasonics Symposium.
3. F. Walls et al, "Measurement of the Short-term Stability of Quartz Crystal Resonators and the Implication for Crystal Oscillator Design", IEEE Trans. on I. and M., Vol. IM-24, No. 1, March, 1975, pp 15-20.
4. S. Eliot et al, "Direct Phase Noise Measurements of SAW Resonators", Proc. 1984 IEEE Ultrasonics Symposium, Oct., 1984, pp 180-185.
5. J. Gagnepain et al, "Excess Noise in Quartz Crystal Resonators", Proc. 37th Frequency Control Sump., USAEC, May, 1983, pp 218.
6. T. Parker, "1/f Frequency Fluctuations in Acoustic and Other Stable Oscillators", Proc. 39th Frequency Control Symposium, May, 1985, pp 97-106.
7. D. Halford, et al, "Flicker Noise of Phase in RF Amplifiers and Frequency Multipliers: Characterization, Cause, and Cure", Proc. 22nd Frequency Control Symposium, USAEC, April, 1968, pp. 340-341.
8. Avantek Semiconductor Device Catalog, 1985, pp. 54-85.

MINIMUM SIDEBAND NOISE IN OSCILLATORS

J.K.A. EVERARD*
King's College London (KQC)
Strand, WC2R 2LS
London
England

SUMMARY

This paper describes how minimum sideband noise can be obtained in resonator type oscillators for any given DC input power. A linear theory is derived which describes how the sideband noise performance in oscillators is dependent on the amplifier noise figure (F), the voltage gain of the amplifier (G), the unloaded quality factor of the resonator (Q_0), the loaded quality factor (Q_L) and the total RF power in the system (PFED). By defining F, G, PFED, Q_L and Q_0 fundamentally and precisely the theory shows that F, G, and Q_L/Q_0 are interdependent. By incorporating the DC input power into the noise equation it is shown that at particular values of Q_L/Q_0 and hence resonator insertion loss that minimum sideband noise can be obtained for any given DC input power. Experimental verification of the theory is also described.

INTRODUCTION

The number of electromagnetic wave communication systems is rapidly increasing, whereas the available frequency spectrum for their transmission is remaining constant. The channel spacing for multi-frequency systems is therefore becoming progressively narrower. This places increasing spectral requirements on the reference oscillator in both the transmitter and the receiver. Modern portable radios must also provide high performance while operating from a small and portable power supply. The oscillator must therefore provide low noise performance with minimum power drain. This paper describes how minimum sideband noise in oscillators can be achieved for any given DC input power.

This paper proposes a new linear theory with experimental verification which shows how the spectrum of a high efficiency oscillator varies with the amplifier noise figure (F), the voltage gain of the amplifier (G), the unloaded quality factor of the resonator (Q_0), the loaded quality factor (Q_L), and the total power in the oscillating system (PFED). By defining the parameters, F, G, PFED, Q_L and hence Q_L/Q_0 fundamentally and precisely the theory shows that F, G, and Q_L/Q_0 are interdependent. It appears that only three references^{1,2,3} mention a relationship between the gain and Q_L/Q_0 and only one³ demonstrates that a noise minimum exists. In the theory presented in this paper, oscillator power is defined as the total RF power in the system and not the available power out of the amplifier nor the power available at the input of the amplifier which appear to be the usual definitions. The power available only exists if the amplifier is presented with a matched load and this is not a required condition for minimum noise. Further if the power is defined as the input power, the gain disappears from the equation causing totally incorrect analysis of the noise minimum. By defining

the power as the total RF power in the system it is possible to obtain an expression which shows how the oscillator sideband noise power varies with DC input power by incorporating the amplifier efficiency. This allows the design of oscillators which have minimum sideband noise for minimum DC input power.

A new highly efficient oscillator configuration, based on the new theory, has been designed which produces minimum sideband noise for minimum DC input power^{4,5}. This configuration also reduces the pulling effect of the load and could therefore be used directly as a transmitter.

Experimental verification of the theory has been obtained at 1 MHz between the limits $0.09 < Q_L/Q_0 < 0.93$. A 150 MHz low noise oscillator, designed using the same approach, has also been successfully demonstrated⁶.

The theory is based on calculating the transfer function of a positive feedback amplifier and shows how the input noise is amplified to produce the output spectrum of an oscillator.

OSCILLATOR THEORY

A model of the high efficiency oscillator is shown in Fig.1.

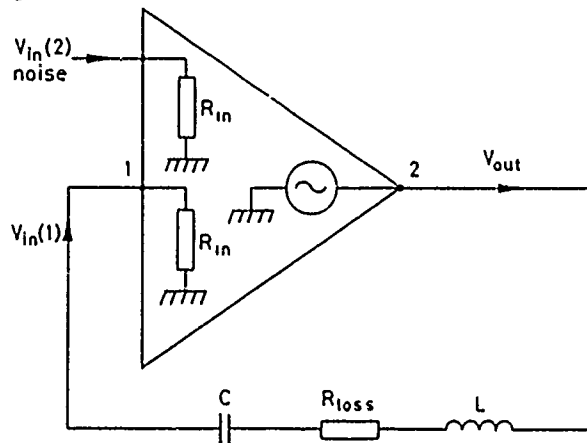


Fig. 1 Model of high-frequency oscillator

This consists of an amplifier with zero output impedance, a known input impedance and a resonant positive feedback network. The zero (low) output impedance of the amplifier is achieved by using a switching output stage which also produces high efficiency by reducing the power dissipated in the amplifier. It also reduces the pulling effect of the load. In this example the resonant network was a series tuned LC circuit with an equivalent series loss resistor which defines the unloaded quality factor, Q_0 .

Input $V_{in(2)}$ was used at the input of the amplifier to model the effect of noise. In a practical circuit the noise would come from the amplifier. The noise voltage $V_{in(2)}$ was assumed to be added at the input and was dependant on the input impedance of the amplifier, the source resistance seen by the amplifier and the noise

*This work was performed at Philips Research Laboratories and is continuing at King's College London (KQC) with sponsorship from the Science and Engineering Research Council (UK). Dr Everard is a lecturer at King's College London (KQC) and a consultant to Philips Research Laboratories.

figure of the amplifier. In this analysis, the noise figure under operating conditions, which takes into account all these parameters was defined as F. This varies from amplifier to amplifier and will be discussed later.

The voltage transfer characteristic is therefore:

$$\frac{V_{out}}{V_{in(2)}} = \frac{G}{1 - \beta G} \quad (1)$$

where G is the voltage gain of the amplifier, β is the voltage feedback coefficient between the nodes 1 and 2 and $V_{in(2)}$ is the input noise voltage.

By evaluating β and assuming $Q_L dF/F_0 \ll 1$ where dF is the offset frequency in Hz and F_0 is the centre frequency in Hz.

$$\frac{V_{out}}{V_{in(2)}} = \frac{G}{1 - GR_{in}} \frac{(1 + 2jQ_L dF/F_0)}{(R_{loss} + R_{in})} \quad (2)$$

$$\text{where } Q_L = 2\pi F_0 L / (R_{loss} + R_{in}) \quad (3)$$

At resonance dF is zero and $V_{out}/V_{in(2)}$ is very large. The output voltage is defined by the maximum swing capability of the amplifier and the input voltage is noise. This is effectively saying that at resonance the amplifier gain is equal to the insertion loss ($GB = 1$). The gain of the amplifier is now fixed by the operating conditions.

The denominator of equation 2 is approximately zero therefore:

$$\frac{V_{out}}{V_{in(2)}} = \frac{G}{2jQ_L dF/F_0} \quad (4)$$

It should be noted that this equation does not apply very close to carrier where the equation would predict that V_{out} is larger than the peak voltage swing of the amplifier, however the equation is accurate in low noise oscillators to frequency offsets considerably closer than 1 Hz to carrier, even for microwave oscillators, as the input noise voltage in a 1Hz bandwidth is usually less than 10^{-9} Volts.

It is necessary to decide where the limiting occurs in the amplifier. In this instance limiting is assumed to occur at the output of the amplifier as this is the point where the maximum power is defined by the power supply. In other words the maximum voltage swing is limited by the power supply. To investigate the ratio of the noise power in a 1 Hz sideband to the total output power the voltage transfer characteristic can now be converted to a characteristic which is proportional to power. This was achieved by investigating the square of the output voltage at the offset frequency (V_{outdf}) and the square of the total output voltage ($V_{outmaxrms}$). Only the power dissipated in the oscillating system and not the power dissipated in the load was of interest. The low output impedance of the amplifier ensured that the load did not affect the noise performance.

The input noise power in a one Hz bandwidth is FkT where kT is the noise power that would have been available at the input had the source impedance been equal to the input impedance (R_{in}). T is the operating temperature and k is Boltzmann's constant. F

is the operating noise figure which includes the amplifier parameters under the oscillating operating conditions such as the source impedance seen by the amplifier. The square of the input voltage was therefore $FkTR_{in}$.

When the total RF feedback power PFED is defined as the power in the oscillating system excluding the losses in the amplifier and most of the power is assumed to be close to carrier then PFED is limited by the maximum voltage swing at the output ($V_{outmaxrms}$) of the amplifier and the value of $R_{loss} + R_{in}$.

$$PFED = \frac{(V_{outmaxrms})^2}{R_{loss} + R_{in}} \quad (5)$$

The ratio of sideband noise power to the total power in the system is therefore:

$$L(fm) = \frac{(V_{outdf})^2}{(V_{outmaxrms})^2} = \frac{GFkT}{8(Q_L)^2 PFED} (F_0/dF)^2 \quad (6)$$

where:

$$PFED = (\text{DC input power to the system}) * \text{EFFICIENCY.}$$

This equation can now be used to obtain minimum sideband noise for any given DC input power.

For minimum noise the noise figure (F) and the value of G/Q_L^2 should be as small as possible. It should be noted however that F, G and Q_L are directly related to each other and thus cannot be varied independently.

The noise equation may be rewritten thus:

$$L(fm) = \frac{FkT}{8(Q_0)^2 (Q_L/Q_0)^2 (1 - Q_L/Q_0) PFED} (F_0/dF)^2 \quad (7)$$

$$\text{where } G = \frac{Q_0 R_{loss}}{Q_L R_{in}} \quad (8)$$

If F is assumed to be constant then the noise equation is minimum when:

$$\frac{dL(fm)}{d(Q_L/Q_0)} = 0 \quad (9)$$

Minimum noise therefore occurs when $Q_L/Q_0 = 2/3$ which sets the voltage gain of the amplifier to be 3.

It should be noted that for these equations to apply it was assumed that F and therefore the input noise voltage was constant and independent of the source impedance presented to the amplifier, i.e. Q_L/Q_0 . Optimum oscillator noise performance can still be met if the amplifier has a minimum noise figure when the source resistance is $2R_{in}$. If optimum noise match occurs far from the point at which the source resistance is $2R_{in}$ then the noise equation should incorporate F as a function of Q_L/Q_0 .

The noise performance degradation with resonator loading may now be examined. Fig.2. shows the noise degradation with Q_L/Q_0 from which it can be seen that as Q_L/Q_0 approaches one (ie the required amplifier gain tends to infinity) or when Q_L/Q_0 tends to zero (ie the amplifier gain tends to 1) the noise performance is markedly degraded.

noise on a spectrum analyser. The variation of sideband noise power with Q_L/Q_0 was measured at 5 KHz offset for an input noise power of -110 dBm/Hz, where the input power was flat from 20 Hz to 50 MHz. The experimental results are shown in Fig.2. and show that minimum sideband noise occurs when Q_L/Q_0 is 2/3 which agrees with the theory. The absolute value of the sideband noise power was found to be 8 dB higher than predicted by theory. This may be due to noise being mixed from other frequencies as the unloaded Q (Q_0) in this experimental oscillator was kept low to enable easy noise measurements. This may well be reduced in a normal oscillator where the unloaded Q (Q_0) would be considerably higher.

The effect of changing Q_0 was checked at the optimum operating conditions by reducing the effective series resistance of the 200 micro-Henry coil (200 Ohms) by a factor of two to 100 Ohms, and comparing the sideband noise when the 100 micro-Henry (100 Ohms) coil was in circuit. In other words Q_L/Q_0 was kept the same but the unloaded Q (Q_0) was increased by a factor of two. The sideband noise power was found to reduce by 5.6 dB. This agrees with the theory ($1/(Q_0)^2$) which predicts a reduction of 6dB.

The effect of altering the total feedback power PFED was checked by reducing the output voltage swing by a factor of two which reduced PFED by 6 dB. The sideband power at 5KHz offset changed by less than 2 dB. This confirmed the theory because the total power was reduced by 6 dB but the sideband power remained approximately constant.

CONCLUSIONS

A linear theory, which describes the optimum operating conditions for minimum sideband noise in oscillators, has been derived and experimentally verified.

The basic requirements for low noise oscillators were found to be:

- 1) A Low noise Figure
- 2) A Large unloaded Q
- 3) A Low output impedance for high efficiency and reduced frequency pulling by the load.
- 4) The ratio of loaded Q to unloaded Q should be 2/3. This may be modified to incorporate the variation of amplifier noise figure with Q_L/Q_0 if this is significant.
- 5) If requirements 1 to 4 cannot be improved upon, then further improvement in noise performance can be achieved by using higher output powers. However care needs to be taken not to overdrive the input, thereby changing the input impedance and noise figure.

Requirement 4 is often missed when oscillators are designed empirically because for most coils a large transforming ratio is usually required. For example at 150 MHz, a coil of value 150 nano-Henries with a Q_0 of 300 has an effective series resistance of less than 1 Ohm. This means that to achieve minimum noise the amplifier input impedance needs to be transformed down to less than 1/2 an Ohm. Many oscillators are probably operating in the region where Q_L/Q_0 is considerably less than 0.1, at which point the noise performance is markedly degraded.

It should be noted that crystals often have a series resistance around 50 to 100 Ohms which is ideally suited to 50 Ohm amplifiers. It is emphasised, however, that the amplifier power level should be kept low to avoid damage and drift to the crystal characteristics.

An amplifier with similar input and output impedances may be used but is likely to be a class A/B amplifier with an efficiency of only a few percent. By using a switching amplifier with low output impedance, a large improvement in noise performance for a given DC input power could be obtained as well as reducing the pulling effect of the load.

REFERENCES

- 1) W.H. Hayward. "Introduction to Radio Frequency Design." Prentice Hall.
- 2) P.A.Moore, S.K.Salmon. "Surface Acoustic Wave Reference Oscillators for UHF and Microwave Generators." IEE.Proc.Vol.130,Pt,H,No.7,Dec.1983 pp.477-482.
- 3) T.E. Parker. "Current Developments in Saw Oscillator Stability." Proceedings of the 31st annual symposium on Frequency control, US Army electronics Command, Fort Monmouth, NJ,1977,pp.359-364.
- 4) J.K.A. Everard. Patent Application 8502565 "Low noise high efficiency oscillator." Feb.1985.
- 5) J.K.A. Everard. "Low noise power efficient oscillators: theory and design". To be published in the next issue of the IEE proceedings part G.

DIGITAL TEMPERATURE COMPENSATION OF CRYSTAL OSCILLATORS USING TEMPERATURE SWITCHES

Željko Aleksić
Dragan Vasiljević

Faculty of Electrical Engineering, University of Belgrade,
Bulevar Revolucije 73, 11 000 Belgrade, Yugoslavia

Aleksandra Pavasović

Institute Mihailo Pupin, Volgina 15, 11 000 Belgrade, Yugoslavia

Summary

The design of low-cost very-low power consumption temperature compensated crystal oscillator (TCXO) having medium frequency to temperature stability is considered in this paper. The described temperature compensation method is based on the application of simple low-power semiconductor temperature switches which are set on various threshold temperatures of the operating range and yield binary output indicating that the preset temperature is reached. Their output is digital and processed by combinational digital circuits (gates and multiplexers) to obtain a preferred varicap diode voltage. In this way, we have the digital compensation method with an easy trimming procedure but without expensive and numerous components (A/D, D/A, PROMS, precision temperature probe). The drawback of our compensation method is that its accuracy is proportional to the number of installed temperature switches. By using twelve switches and a quartz crystal with temperature stability of 40ppm in the range from -40°C to $+70^{\circ}\text{C}$, we obtained TCXO frequency stability of ± 2 ppm.

Introduction

The temperature compensated crystal oscillators (TCXO) are widely used in electronic industry for data processing and transmission. In general, TCXO should be low-cost, small sized, with low power consumption and a prescribed frequency accuracy.

Most of the currently produced TCXO's are compensated using analog techniques achieving frequency stabilities up to 0.5 ppm in the temperature range from -40°C to $+70^{\circ}\text{C}$. Fundamental problems appearing in analog compensation network design are due to small component value tolerances and interdependence of network adjustment on various temperatures, leading to cheap network components and expensive trimming procedures^{1,2}.

Some improvement may be obtained by using a computer for network parameter calculations and/or analog generator of voltage-temperature characteristic in order to realize an independent segment adjustment. Unfortunately, some difficulties remain: a lack of accurate network parameter values necessary for calculations and the lack of true independence of segments in analog generator output voltage^{3,4,5,6}.

These problems can be resolved by a digital method of crystal oscillator temperature compensation (DTCXO) which is practical for frequency stabilities better than 0.5 Hz/MHz. Certainly, this method of compensation includes a large number of components and has a major limitation due to the quartz crystal unit thermal

hysteresis^{7,8}.

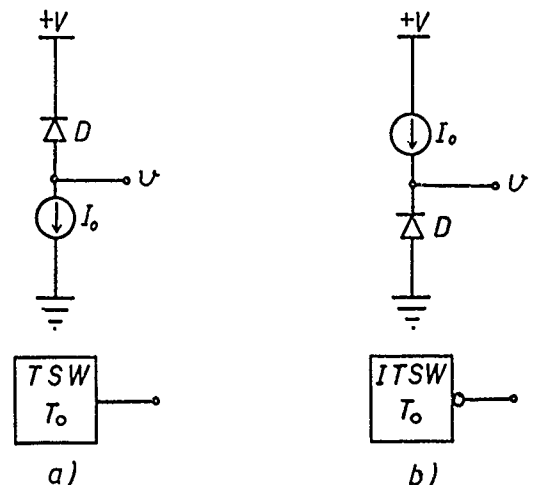
Our goal was to design a simple, cheap, low-power and easily adjustable compensation network for a moderate frequency stability TCXO circuits. We opt for digital compensation in order to obtain independent trimming on various temperatures, but our concept is different from the classical DTCXO. Instead of a temperature transducer, A/D converter, PROM and D/A converter, we use a sufficient number of temperature switches, a coder and analog switches to produce analog voltage applied to the varicap diode.

It is possible even to eliminate the varicap diode (in this case the temperature switches are used for connecting capacitor network to the oscillator circuit, similar to the solution given in ref.9). The main drawback of our compensation method is that the number of installed switches increases with required frequency accuracy.

The principle of temperature compensation based on temperature switches

Temperature switch

A temperature switch provides a high digital output when ambient temperature T_a exceeds the preset value T_0 , and low output for $T_a < T_0$ (noninverting switch). It is also possible to design an inverting temperature switch with low output for $T_a > T_0$ and high output for $T_a < T_0$. The temperature switch consists of a series connection of a constant current source (whose current I_0 is independent from the temperature and the reverse biased diode (whose reverse current I_s is used for temperature sensing (Fig. 1)¹⁰.



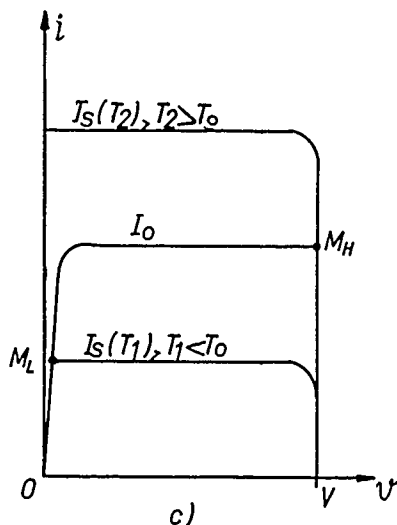


Figure 1 Noninverting temperature switch (a), Inverting temperature switch (b), Illustration of noninverting switch operation (c).

A temperature controlled voltage generator based on temperature switches

The typical voltage to temperature characteristic necessary for the varactor diode compensation of oscillator frequency drift in the operating temperature range, is shown by a solid line on Fig. 2. The digitally generated curve is shown on the same figure by a dashed line. The number N of voltage jumps depends on

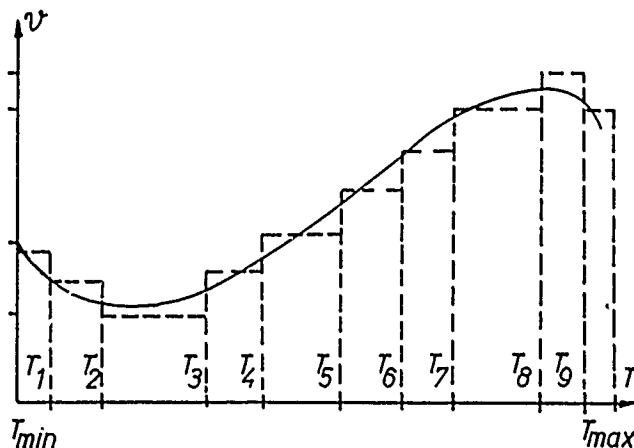


Figure 2 Varactor voltage to temperature characteristic: preferred curve (solid line), digitally realised curve (dashed line).

crystal and oscillator circuit maximal frequency to temperature drift ($f_{max} - f_{min}$) and frequencies on the ends of temperature range ($f(T_{min})$ and $f(T_{max})$). Also, N is determined by a desired frequency to temperature compensation accuracy E given in part per million. So, we have:

$$N = \frac{|f(T_{min}) - f_{min}| + |f_{max} - f_{min}| + |f(T_{max}) - f_{max}|}{E} \quad (1)$$

The temperature controlled voltage generator is realised using one temperature switch for each temperature T_i when voltage on Fig 2 jumps, a coder and a D/A converter. To obtain high simplicity, the coder is realised by wiring and the D/A converter by analog

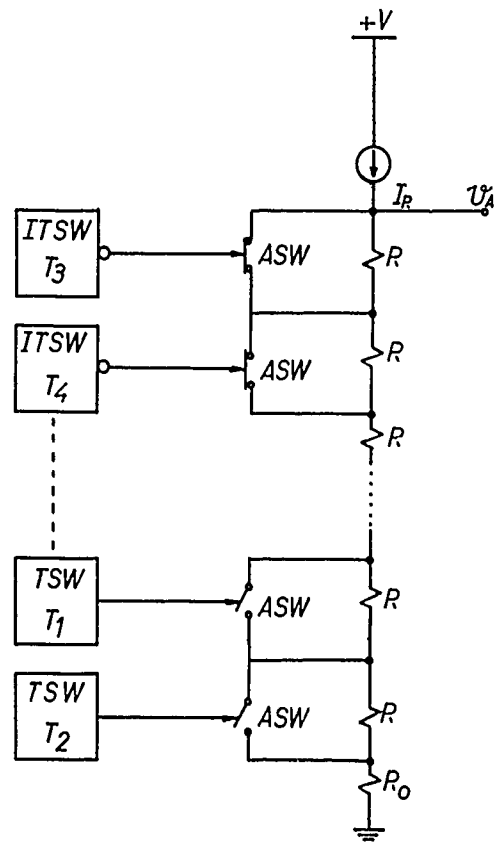


Figure 3 Voltage to temperature generator based on temperature switches

switches (ASW) and a resistive network (Fig.3). The inverting temperature switches are used for a voltage increase and noninverting temperature switches for voltage decrease.

The voltage generator set-up process

The set-up starts from the lowest temperature point. After an initial frequency trimming by R_0 on Fig. 3, the temperature of the oscillator surroundings is raised until the deviation from the nominal frequency is equal to the prescribed value E . This temperature is the first boundary temperature point where temperature switch must be activated by adjusting the current of its current source. The change of the output of a temperature switch opens or closes the analog switch; the later causes a jump up or down in output voltage, which changes oscillator frequency in the opposite direction amounting to $2E$. This treatment is repeated up to the highest operating temperature.

The proposed solution provides a simple trimming procedure, providing the highly independent adjusting in all temperature intervals. The characteristics of the components used (temperature sensing PN junctions, crystal unit and oscillator circuit) can be previously undetermined, but should lie in the predetermined ranges of values.

Experimental results

The temperature-dependent voltage generator is realised as shown on Fig. 3. The set-up procedure was realized in one temperature cycle as described above. The realisation of the noninverting and inverting

temperature switches is shown on Fig. 4. Transistor TR_1 amplifies the inverse diode current and increases the output resistance of the switch. This transistor can be reverse biased to obtain lower current gain.

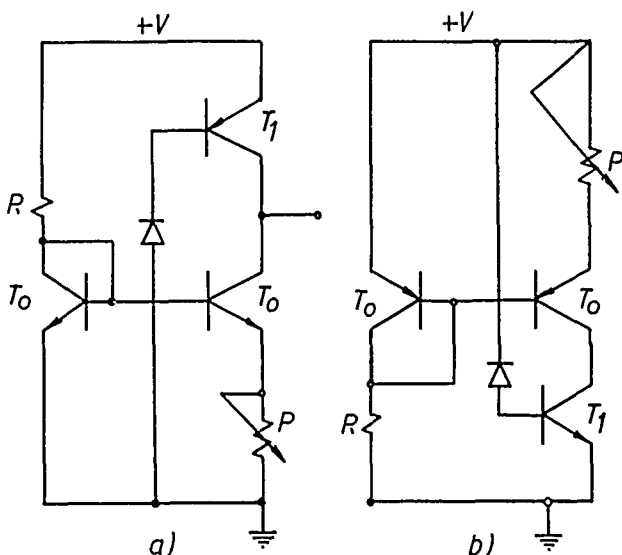


Figure 4 Realisation of the noninverting (a) and inverting (b) temperature switches. Used components: fast recovery diodes and general purpose transistors.

The standard CMOS analog switches (4066) are used for 10 000 Ohm resistors R switching (Fig. 3). The 10 MHz uncompensated oscillator was used with frequency to temperature variations from: $f(T_{\min}) = f_{\min} + 9$ ppm, to $f(T_{\max}) = f_{\max} = f_{\min} + 40$ ppm. Using two non-inverting and ten inverting temperature switches, we obtain a temperature compensated oscillator with accuracy of ± 2 ppm in temperature range from -40°C to $+70^\circ\text{C}$ (Fig. 5).

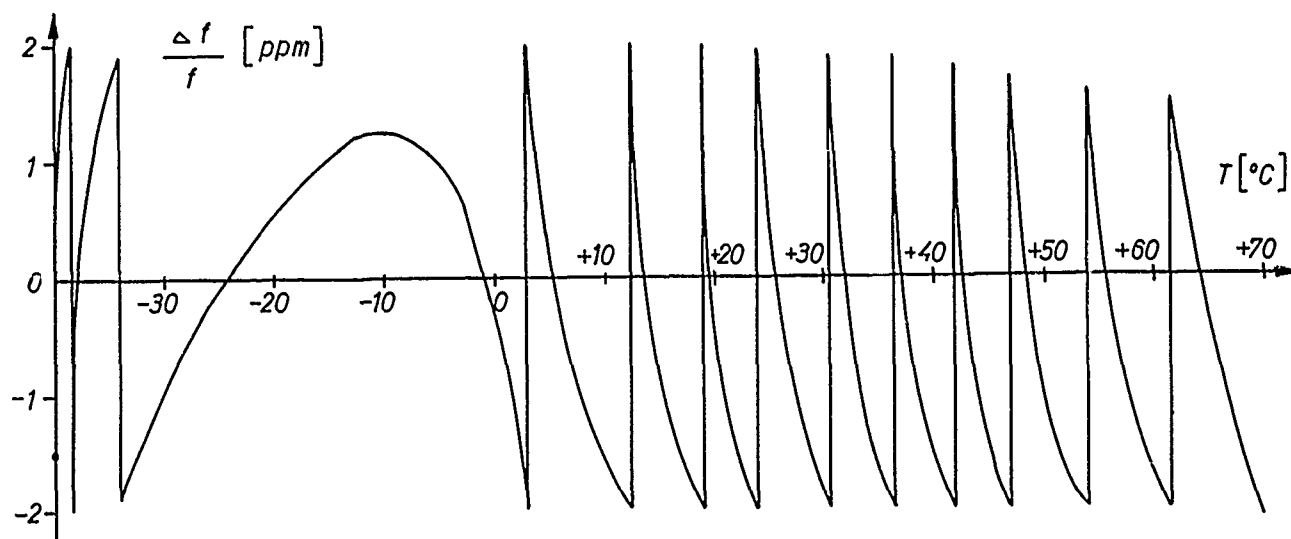


Figure 5 Frequency to temperature characteristic of an oscillator compensated by new digital technique.

The current consumption of compensation network is lower than 0.1mA when ambient temperature is different from any threshold temperature T_i defined on Fig. 2. When the ambient temperature reaches threshold temperature T_i , the corresponding switch gives output voltage

which can drive the CMOS analog switch input inverter in regime where both transistors conduct. This rises the supply current to few milliamperes. This current jump can be lowered using ternary CMOS inverters¹¹. A more efficient way is to insert a constant current source between drains of P channel and N channel transistor of the CMOS input inverter. In this way we obtain an additional supply current increase over 0.1 mA only in the range of 0.01°C about threshold temperature. This gives a very small possibility (0.1%) that supply current of the compensation circuit exceeds 0.2mA.

Conclusion

The described method for digital temperature compensation has three main advantages: simple and independent compensation circuit adjustment in one temperature cycle; low component price; and low power consumption.

The main disadvantage of this technic is a large number of temperature switches, needed for high precision temperature compensation.

Two basic technical problems in the realisation of the proposed compensation method are: a large number of transistors and a low inverse current of the diode on low temperatures. They can be resolved by an integrated circuit design and a special PN junction. The shown solution is only an experimental confirmation of the described compensation method and its realisation must be further improved for a direct practical implementation.

Acknowledgements

The authors wish to thank Bojan Dobnikar, Director of Crystal Division, IMP, for his encouragement. They also acknowledge Milan Stojanović and Željko Tepavac for their help in the experimental procedures.

References

1. —, "Temperature compensated crystal oscillator panel discussion", Proc. of the 36th Annual Freq.

- Cont.Symp., pp. 562-563, 1982.
2. Frerking M.E., "Methods of temperature compensation", Ibid, pp. 564-570.
 3. Vovell P.G., "Recent improvements to TCXO", Proc. of the 22th Annual Freq. Con. Symp., pp. 311-324, 1968.
 4. Newell D.E. and Hinnah H., "Automatic compensation equipment for TCXO'S", Ibid, pp. 298-311.
 5. Newell D.E. and Hinnah H., "A report on TCXO'S and segmented compensation", Proc. of the 23th Annual Freq. Cont. Symp., pp. 187-191, 1969.
 6. Wilson J.S., "An improved method of temperature compensation of crystal oscillators", Proc. of the 37th Annual Freq. Cont. Symp., pp. 442-447, 1983.
 7. Renard A.M. and Barnhill K., "Digital temperature compensated oscillators", Proc. of the 35th Annual Freq. Cont. Symp., pp. 455-457, 1981.
 8. Pollard G.B., "A digitally compensated hybrid crystal oscillator", Proc. of the 36th Annual Freq. Cont. Symp., pp. 474-479, 1982.
 9. Uno T. and Shimoda Y., "A new digital TCXO circuit using a capacitor-switch array", Proc. of the 37th Annual Freq. Cont. Symp., pp. 437-441, 1983.
 10. Aleksić Z.J. and Vasiljević D.M., "Low power temperature detector and rate of temperature change detector", IEEE Trans.Instr. Meas. (to be published).
 11. Koanantakool H.T., "Implementation of ternary identity cells using C.M.O.S. integrated circuits", Electronics Letters, Vol. 14, No. 15, pp. 462-464, 20th July, 1978.

A CRYSTAL OSCILLATOR WITH BIDIRECTIONAL
FREQUENCY CONTROL AND FEEDBACK ALC

A. Benjaminson

Systematics General Corporation

Arlington, Virginia

Summary

Continued work on bridge oscillators has shown that the four-arm bridge configuration provides a greater pulling range than do conventional oscillators. The analysis reviewed herein shows that the oscillator frequency can be pulled both above and below the series-resonant frequency without the use of inductors. The additional range can be obtained by locating a trimming capacitor in an arm of the bridge not occupied by the resonator.

The effect of value changes in all four arms are discussed, plus the effect of sustaining transistor feedback capacity on the frequency of operation.

The equations necessary to complete understanding of this technique are derived and discussed in the paper.

The bridge circuit also simplifies and improves the application of ALC to crystal oscillators. A JFET can be placed in the bridge arm adjacent to that containing the resonator to provide automatic compensation for changes in crystal resistance with temperature, frequency offset, and replacement..

A computer program has been developed to provide a complete circuit design with component values and operating parameters.

Background

The bridge oscillator was first conceived by L.A. Meacham¹ in 1938. At the time it was hailed as "providing the greatest frequency stability of any vacuum tube oscillator yet devised"^{2,3} and was used in the frequency standards of the Bell System, the National Bureau of Standards, and the British Post Office. Edson, in his classic text "Vacuum Tube Oscillators"², showed that "the balancing action of the bridge tends to increase the effective Q of the series resonant circuit (crystal) by magnifying the phase shift produced as a result of any frequency deviation"

The work of Pierce and Colpitts had already been published and their oscillators were in general use, yet the bridge oscillator with its proven merits soon faded from sight until the era of solid state amplifiers sparked an occasional return. The reasons for the early demise of the original tube circuits were clear, the high tube capacities, and poor transformer material made parasitic oscillations a serious problem as crystal frequencies were increased to 1 MHz and higher. Why, with wide bandwidth amplifier designs available, the bridge oscillator did not catch on is a mystery, that has not been explained.

Now that the SC-cut crystal is being used more often in place of the AT and BT cuts, the bridge oscillator should be considered as a serious candidate where low noise and simple overtone and B-mode rejection circuits are required. Being a series-mode crystal oscillator, the frequency can be pulled above the series resonant frequency over a greater range than antiresonant-mode oscillators can. In this paper the author will discuss a technique for pulling the frequency below series resonance as well, effectively tripling the available pulling range, and thereby reducing the fabrication tolerance for the crystal manufacturer, all without the use of inductors.

Bridge Oscillator Configuration

The solid-state bridge oscillator, unlike Meacham's original, usually uses a differential amplifier configuration instead of transformers to couple the bridge to the sustaining amplifier.⁴ This provides a high input impedance, minimizing the load on the output of the bridge without the use of a tuned transformer. The simplest circuit possible is shown in Figure 1, where a differential RF amplifier such as the CA 3001 also provides a low output impedance for driving the bridge circuit.

The necessary condition for oscillation can be described in the terms used for analysis of feedback circuits, where a positive feedback factor β_p can be defined as

$$\beta_p = Z_3 / (Z_3 + Z_4) \quad (1)$$

and a negative feedback factor β_n as

$$\beta_n = Z_1 / (Z_1 + Z_2) \quad (2)$$

The gain, A_o of the sustaining amplifier can be considered first as being modified by negative feedback only, so that

$$A_n = A_o / (1 + A_o \beta_n) \quad (3)$$

When positive feedback is applied to the stabilized amplifier with gain A_n , oscillation will occur when

$$\beta_p A_n = 1 \quad (4)$$

Substituting eqn (3) into (4) the result reduces to

$$A_o (\beta_p - \beta_n) = 1 \quad (5)$$

(The implication is that the terms of eqn. (5) are all real, however complex values can be used with-

out error provided the product is real and equal to one.)

This has been covered in greater detail in the author's previous papers on the subject ^{4,5} but without sufficient emphasis on the importance of crystal resonator location in the circuit.

An examination of eqn. (5) alone suggests that a series resonator can be located either in position Z_4 or Z_1 . In the first position, the low resistance at resonance can be used to raise β_p enough to satisfy the conditions for oscillation. (This assumes fixed resistances in the other three locations.)

If the resonator is placed in the second position (Z_1) it can then be used to reduce β_n enough to also satisfy the oscillatory condition. Which location is best? Although most published circuits seem to favor the former position (Z_4) for the resonator, the author contends that the Z_1 position is best. An examination of eqn. (3) indicates that if the resonator is in the Z_1 position, the gain of the sustaining amplifier (without positive feedback) will approach unity at all frequencies above and below the bandwidth of the resonator, and will act as a sharply-tuned amplifier in the region of resonance.

This results in lowered sideband noise in the oscillator's spectrum since the thermal and shot noise generated in the input circuit, are not amplified outside the resonator bandwidth.

Locating the resonator in the Z_4 position, on the other hand, does not permit any reduction in amplifier gain outside the resonator bandwidth, since the gain is fixed by the constant impedance ratio of Z_1+Z_2/Z_1 .

A second advantage of the Z_1 position is that one side of the crystal is grounded. This reduces the triad of capacitances around a floating crystal connection to a single reduced capacity across the crystal to ground. This not only simplifies the circuit analysis, but in applications where the crystal is used as a transducer for pressure, temperature, etc. it also simplifies the shielding arrangement.

Although the circuit in Figure 1 uses resistors to provide positive feedback, any pair of impedances can be used as long as the phase shift is zero at the operating frequency. (A resistor is essential in the Z_2 position for operation at series-resonance.) This freedom is advantageous when overtone crystals are used, or when the B-mode of an SC-cut crystal must be rejected. As shown in Figure 2, a capacitive divider is used to provide positive feedback, and is shunted by an inductance to tune the circuit to the mode frequency of interest. The inductance may also be used as a coupling medium to a low-impedance secondary winding for driving a transmission line or buffer amplifier.

The use of resistors to establish positive feedback has the potential for causing changes in frequency with changes in shunt capacity across the Z_3 , Z_4 elements. Capacity changes here, particularly those due to the Miller Effect induce phase, and therefore, frequency changes.

This is not true of reactive elements, since only changes in the amplitude of β_p are possible which will influence the level of signal only.

When a differential amplifier with two open collectors is used the tuned circuit is essential, and can be used either single-ended, as shown in Figure 2, or placed between the collectors as is done in the full differential crystal oscillator. (Figure 5)

Frequency Adjustment in the Bridge Oscillator

The circuit in Figure 2 will operate at the series resonant frequency of the crystal, assuming no phase shift in the amplifier or tuned circuit, since Z_1 must be resistive to match the zero phase angle of Z_2 .

The frequency can be adjusted on either side of series-resonance by adding capacitors C_1 and/or C_2 to the circuit as indicated in Figure 3.

Considering only the effect of C_1 initially, its capacitive reactance will raise the frequency until the resonator produces a cancelling inductive reactance to retain the zero degree phase angle required of Z_1 . The action is the same as that of a load capacitor in antiresonant circuits, except that there is no upper limit on the size of C_1 . There is a lower limit though, imposed by the shunt capacity C_0 . The offset frequency produced by C_1 alone will be

$$df/f = C_1/2C_1 \quad (6)$$

assuming C_1 is much larger than C_0 . (C_1 is the motional capacitance of the crystal).

Returning to Figure 3, consider a finite value for C_2 , the capacitance in series with R_2 . This will introduce a leading angle in Z_2 , requiring a similar matching angle in Z_1 , i.e.

$$\arctan 1/\omega C_1 R_1 = \arctan 1/\omega C_2 R_2 \quad (7)$$

$$\text{or } C_1 R_1 = C_2 R_2$$

As shown in the detailed analysis of Appendix I, the offset frequency becomes

$$df/f = -(C_1/2)(R_S/R_2 C_2) \quad (8)$$

The resulting control of df allows the offset frequency to be adjusted both above f_s (by C_1) and below f_s (by C_2).

Since there is always some feedback capacitance between the output and input of the sustaining amplifier, this is represented by C_f in Figure 3, and has been evaluated in Appendix I. As shown in equations (8a), (8c) in the appendix, C_f will also lower the frequency by contributing to the effect of C_2 , thus

$$df/f = (C/2)(1/C_1 - R_S/R_2)(1/C_2 + C_f/C_2^2) - R_2 R_S \omega^2 C_f \quad (9)$$

Although there may also be feedback capacity to the positive input terminal, this can be absorbed by the capacitors in the tuned circuit shown in Figure 2, and need not influence the phase conditions.

Bidirectional frequency control can be implemented by using two separate trimmers for C_1 and C_2 , or by using a single differential trimmer. In the first arrangement C_1 can be used to raise the frequency with C_2 set at maximum, or C_2 can be used to

lower the frequency with C_1 set at its maximum value.

A differential trimmer is an elegant method, since it combines C_1 and C_2 in such a manner that one value decreases while the other increases. Figure 4 shows a curve calculated for such an arrangement, assuming each capacitor has a range of 5 to 25 pfd., and that $R_s = R_2 = 500$ ohms. C_f , being only 1 pfd., has little effect on the curve.

If R_2 is less than R_s , which is usually the case for designs where R_s is small, then, with equal value capacitors, the negative excursion of the curve will be decreased, in proportion to R_s/R_2 . This can be avoided by scaling C_2 down by the same ratio as R_s/R_2 .

Automated Level Control in Bridge Oscillator

ALC is essential for stable oscillator designs, since it controls and stabilizes the current in the resonator, and restricts the signal amplitude to the linear range of the amplifier. In conventional parallel resonant oscillators it is usually applied to the sustaining amplifier. A dc voltage derived by rectification of the oscillator signal is usually applied to a current control element of the amplifier. In this way the transconductance, and thus the gain of the amplifier, is adjusted to maintain a preset value of crystal current.

There are two disadvantages to this technique, the first is that it is difficult to maintain the optimum dc current in the transistor that will produce the best noise figure, and the second is that negative feedback, in the form of an unbypassed emitter resistor, cannot be used, since it inhibits gain control. Such a resistor has been shown to be important in reducing the effects of flicker noise.

The ALC system currently used in the differential crystal oscillator, and applicable to bridge oscillators in general is detailed in Figure 4. This approach does not change the gain of the active device, but controls the negative feedback through the resonator. A JFET is used as a variable resistor that adjusts the value of β_n to satisfy the condition for oscillation discussed earlier. ALC applied in this manner does not degrade the amplifier but permits it to be designed for maximum signal to noise ratio, and a maximum output level. Feedback ALC also provides automatic tracking of changes in crystal resistance due to aging, temperature, or production tolerances. If the apparent resistance of the crystal changes due to frequency pulling, the JFET will automatically track and compensate with a small change in output level.

The ALC control voltage is usually obtained from a voltage-doubler circuit using Schottky diodes for high efficiency. A typical voltage level is between 3 to 6 VDC. The gain of the amplifier between oscillator and rectifier determines the signal level at the oscillator and hence the crystal current. The type of JFET favored is the R' amplifier type as exemplified by the 2N5397. A level-setting control is required because of the wide variation of pinch-off voltage in JFET's.

Acknowledgements

The author wishes to acknowledge the support of the U.S. Army's Electronic Technology and Devices

Laboratory under Contract No. DAAL01-85-C-0445, and the support of my colleague, Paul J. Toste in developing and testing the computer programs involved.

References

- (1) L.A. Meacham, "Bridge-Stabilized Oscillator", Proceedings of the IRE 26(10) pp.1278-1294 (1938).
- (2) W.A. Edson, "Vacuum Tube Oscillators", John Wiley & Sons, 1953.
- (3) "Handbook of Piezoelectric Crystals for Radio Equipment Designers", WADC Technical Report 56-156.
- (4) A. Benjaminson, "Balanced Feedback Oscillators", Proceedings of the 38th Annual Symposium on Frequency Control, June 1984 pp. 327-333.
- (5) A. Benjaminson, "Results of the Continued Development of the Differential Crystal Oscillator", Proceedings of the 39th Annual Symposium on Frequency Control, May 1985 pp. 171-175.

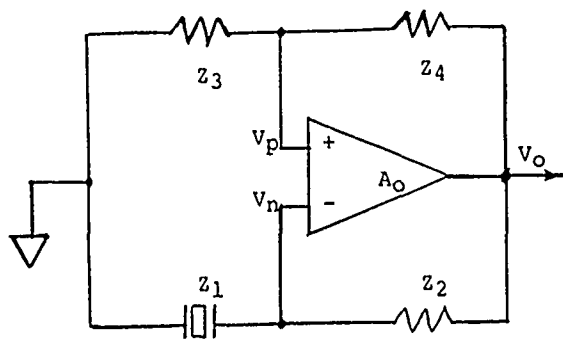


Figure 1. Bridge Oscillator
Basic Configuration

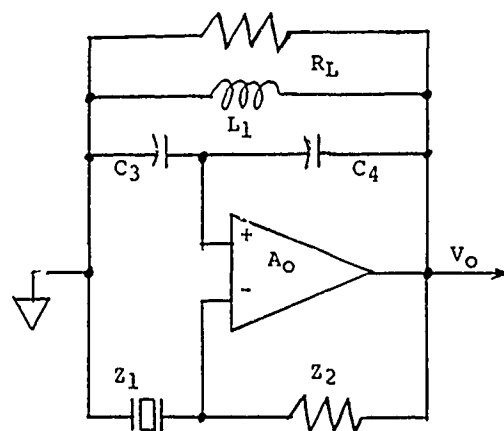


Figure 2. Tuned Bridge Oscillator

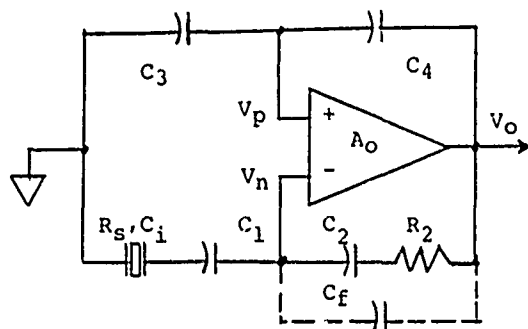


Figure 3. Frequency Control
of Bridge Oscillator

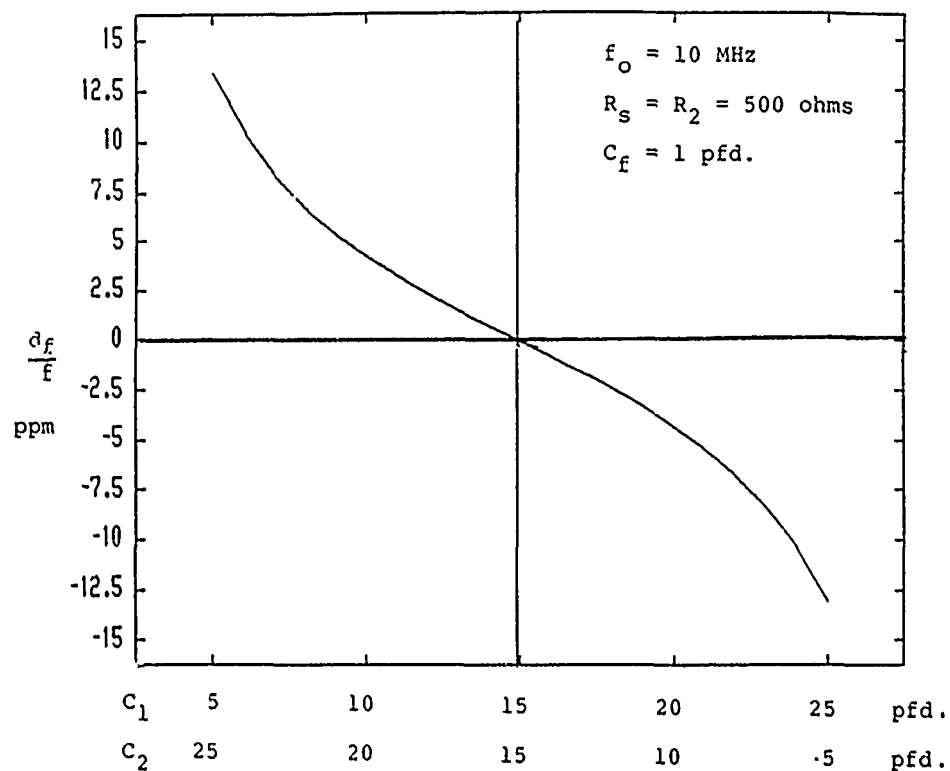


Figure 4. Differential Frequency Pulling

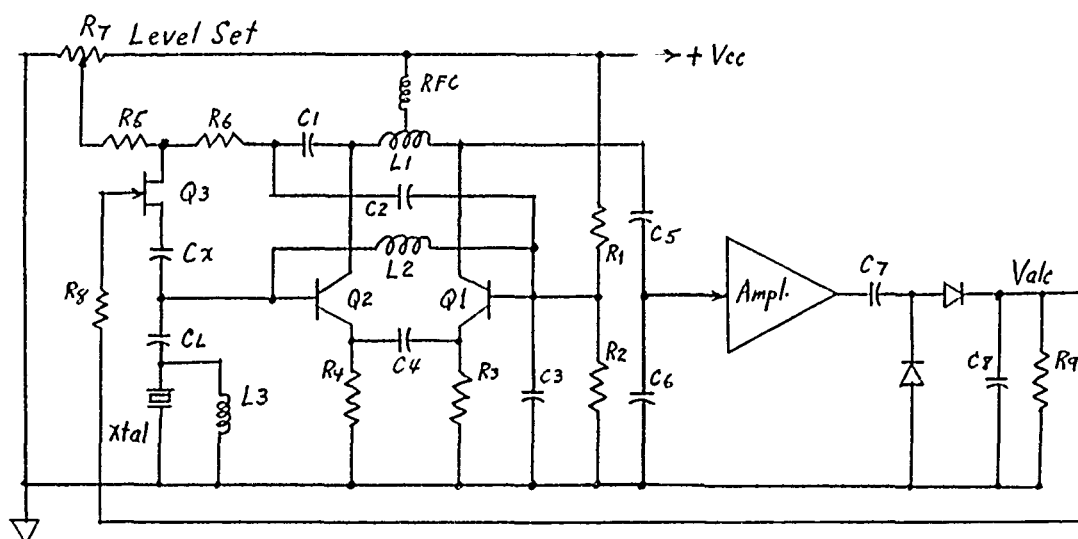


Figure 5. Differential Crystal Oscillator with FET Level Control

Appendix I

AN ANALYSIS OF THE FREQUENCY CONTROL CAPABILITY OF THE BRIDGE OSCILLATOR

Initial Conditions: v_p in phase with v_o ; zero phase shift thru A_o , C_1 is internal feedback capacity of active device, A_o .

Problem: Determine frequency offset from series resonance of crystal (Z_s) as a function of C_1 , C_2 , R_2 , R_s and C_f .

1) Determining the parallel equivalent of C_2 , C_f , R_2 , and calculating the phase angle of the impedance.

$$R_p = R_2 + \frac{X_2^2}{R_2} \quad , \quad \frac{1}{X_p} = \frac{1}{X_f} + \frac{X_2}{R_2^2 + X_2^2} \quad (1)$$

$$\tan^{-1} \phi_2 = \frac{R_p}{X_p} + \frac{R_2^2 + X_2^2}{R_2} \left(\frac{1}{X_f} + \frac{X_2}{R_2^2 + X_2^2} \right) \text{ simplifying}$$

$$\tan^{-1} \phi_2 = \frac{R_2^2 + X_2^2}{R_2 X_f} + \frac{X_2}{R_2} \quad (2)$$

$$2) \text{ Then } \tan^{-1} \phi_1 = \frac{X_1 - X_s}{R_s} \quad (3)$$

3) The conditions for oscillation require

$$A_o (\beta_p - \beta_n) = 1, \text{ or } A_o \left(\frac{v_p}{v_o} - \frac{v_n}{v_o} \right) = 1, \quad (4)$$

so that v_n must be in phase with v_p , therefore

$$\tan^{-1} \phi_1 = \tan^{-1} \phi_2, \text{ or}$$

$$\frac{X_1 - X_s}{R_s} = \frac{R_2^2 + X_2^2}{R_2 X_f} + \frac{X_2}{R_2} \quad (5a)$$

$$\frac{X_s}{R_s} = \frac{X_1}{R_s} - \frac{R_2}{X_f} - \frac{X_2^2}{R_2 X_f} - \frac{X_2}{R_2} \quad (5b)$$

$$\text{But } \frac{df}{f} = \frac{X_s}{2X_i}, \text{ where } X_i \text{ is the reactance} \quad (6)$$

of the motional capacitance of the crystal, subst'g

$$\frac{df}{f} = \frac{1}{2X_i} \left(X_1 - \frac{R_2 R_s}{X_f} - \frac{R_s X_2^2}{R_2 X_f} - \frac{R_s X_2}{R_2} \right) \quad (7)$$

and in terms of capacitance

$$\frac{df}{f} = \frac{C_1}{2} \left(\frac{1}{C_1} - R_2 R_s \omega^2 C_f - \frac{R_s C_1}{R_2 C_2^2} - \frac{R_s}{C_2 R_2} \right) \quad (8)$$

$$\text{Also } \frac{df}{f} = \frac{C_i}{2} \left(\frac{1}{C_1} - \frac{R_s}{R_2} \left(\frac{1}{C_2} + \frac{C_f}{C_2^2} \right) - R_2 R_s \omega^2 C_f \right) \quad (8a)$$

If C_f is neglected then

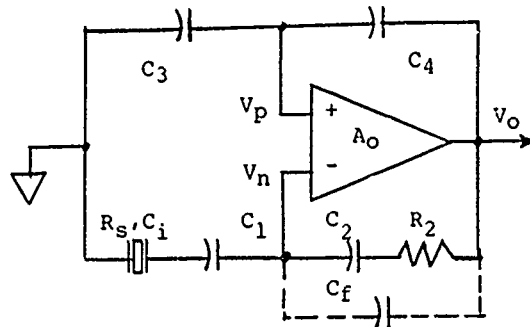
$$\frac{df}{f} = \frac{C_i}{2} \left(\frac{1}{C_1} - \frac{R_s}{R_2} \frac{1}{C_2} \right) \quad (8b)$$

resulting in C_1 controlling an increase in the offset frequency while C_2 controls a decrease in frequency.

If C_1 and C_2 are very large then

$$(5) \quad \frac{df}{f} = \frac{-\omega C_i}{2} \cdot \omega C_f R_2 R_s \quad (8c)$$

and C_f causes a small initial decrease in frequency.



INTEGRATED OVEN CONTROLLED QUARTZ CRYSTAL OSCILLATOR

G. Marianneau, D. Hauden, J.J. Gagnepain

Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté - Besançon
32 avenue de l'Observatoire - 25000 Besançon - France

Summary

A low power, integrated and ovenized, quartz crystal oscillator is presented in this paper. The characteristics of miniaturized quartz resonators vibrating in the 2.5 MHz - 10 MHz frequency range with AT, GT or ZT-cuts are evaluated. The loop amplifier, the electronics of the temperature controlled oven, the temperature probe and the heating power transistors are integrated in an analog gate array circuit. The integrated circuit is mounted together with the resonator and a varactor diode in a small standard oscillator case. This oscillator was primarily designed for buoys of the SARSAT program. SARSAT is a rescue localizing system with both seaborne and airborne applications. The performances of the oscillator are measured in term of short term stability, warm-up time, and power consumption. The goal is to obtain stabilities better than 10^{-9} with power consumptions of the order of 600 mW at -40°C and 200 mW at room temperature. These results are presented, and compared to SARSAT specifications.

Introduction

SARSAT is a satellite system for the search and rescue of planes, boats, or ground mobiles. The SARSAT program organisation relies on three elements. The embarked equipments on satellites NOAA assume reception, preprocessing and distribution of the messages of distress. The reception stations on the ground perform signal processing in order to localize the mobile in difficulties in quasi-real time for the local covering and in delayed time for the main station. The rescue buoys embarked on mobiles must enable, by means of Doppler measurements, identification and localization. The buoys are driven in frequency by means of a stable oscillator (OVS) which defines the performances of the system and mainly the localization accuracy.

The casualties of the SARSAT program need oscillators having a low electrical consumption (400 mW at -40°C) small sizes and a good frequency stability. They must be usable in a wide temperature range with gradients and a low cost is also an important parameter.

A miniaturized oven controlled quartz crystal oscillator was studied and is described in this paper. Miniaturized quartz resonators are associated with an analog gate array which integrates the different electronics functions. Both are mounted in a small standard case and the thermal transfer surfaces are minimized to achieve to SARSAT requirements.

Miniaturized ovenized quartz oscillator

The oscillator is constituted of three parts associated on a hybrid circuit : a small quartz resonator (similar to those used in watches) is connected with an analog gate array specially designed to have the different electronics functions (oscillator loop and output amplifier, temperature probe, oven control and heater) integrated together. Few complementary components : a varactor diode to set the frequency, and adjustable resistances are mounted on the same hybrid circuit.

Three different quartz resonators have been used in the oscillator with the same gate array. The resonant frequency was in the range of 2.5-10 MHz (Fig. 1) :

resonator COMODOR of Dryan Fordahl
resonator CX-MV of ETA Company
and ZT-cut resonator of CSEM and ETA Company

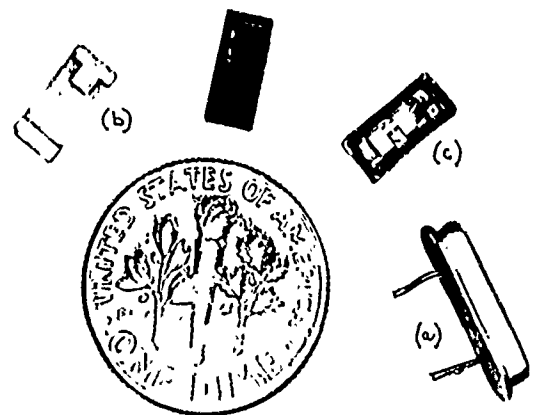


Fig. 1 - Miniaturized quartz resonators : resonator COMODOR (a), resonator CX-MV (b) and ZT-cut resonator (c)

The AT-cut COMODOR resonator is a rectangular AT-cut crystal, elongated along the Z' axis¹⁻⁴ vibrating at 5.075 MHz. It was specially designed for controlled-oven oscillators has a case of the type called bath-tub with a length of 13.3 mm, a width of 3.4 mm and a height of 2.9 mm. The main characteristics are : Q-factor of 350 000, series resistance of 50 Ω , aging of 2×10^{-9} /day and g-sensitivity of 2×10^{-9} /g.

The CX-MV quartz resonator is 10.13 MHz, AT-cut, rectangular crystal manufactured with a photolithographic process and chemical milling⁵⁻⁶ (similar to the process used for watch resonators vibrating at lower frequency. Its rectangular case is well adapted for miniaturized systems. The sizes are $8.38 \times 3.94 \times 2.03$ mm. The Q-factor is 25 000 and the series resistance is 150 Ω .

A miniaturized ZT-cut resonator⁷ developed recently with a photo-lithographic process has a Q-factor of 500 000, serie resistance of 100 Ω , aging of 10^{-8} /day and accelerometer sensitivity equal to 3×10^{-10} /g.

Fig. 2 shows the frequency-temperature characteristics of the three resonators.

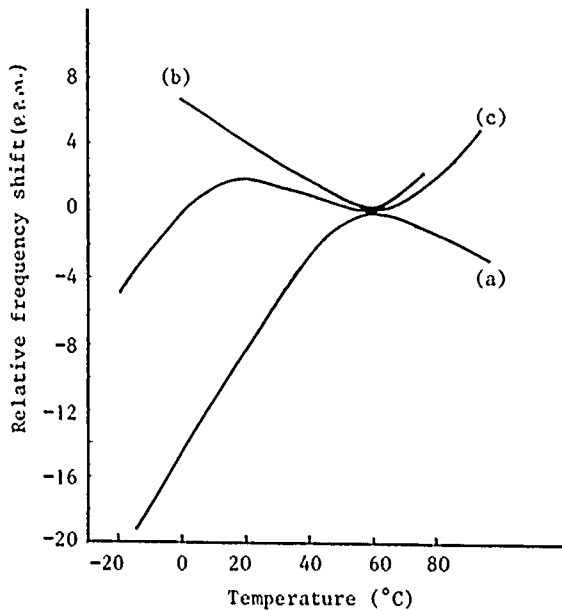


Fig. 2 - Resonator frequency-temperature characteristics. (a) resonator COMODOR, (b) resonator CX-MV, (c) ZT-cut resonator

Figure 3 shows the integrated circuit. It contains the electronic functions of the oscillating loop (modified Buttlar type) and its output amplifier, a voltage regulator, the temperature probe made with a band-gap circuit, the temperature controlled oven electronics and the heater which is constituted by two power transistors integrated in the gate-array.

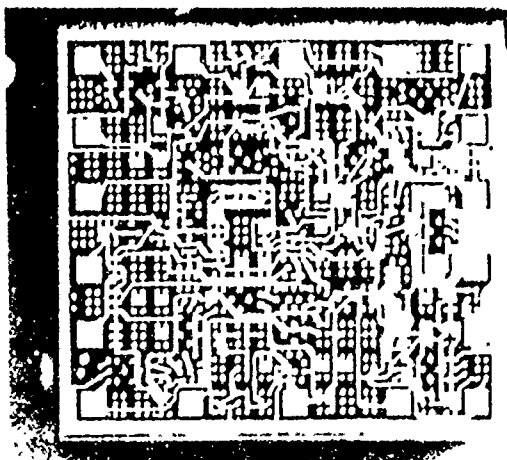


Fig. 3 - Photography of the integrated gate-array circuit

This circuit, when the buoy is operating, is supplied by a voltage which can vary from 20 V to 10 V, and the available power is up to 7 W. It has been manufactured by ATAC Diffusion (France) from an Interdesign analog gate array (MOH). The resonator and the integrated circuit are connected together in a hybrid circuit with the varactor diode and the resistances used to adjust the reference temperature at the turn-over point and the maximum current of the heater.

The oscillators for the SARSAT program need also an output circuit (driven by a pulsed power supply), which is mounted on a second hybrid circuit. All together are clamped in a case that sizes are $36 \times 27 \times 20$ mm. Figure 4 shows the oscillator circuit.

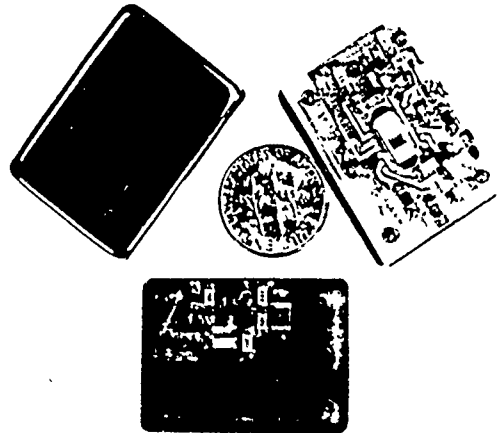


Fig. 4 - Miniaturized ovenized quartz oscillator

Oven controlled oscillator characteristics

The oscillator is characterized by its short term frequency stability, power consumption, external temperature sensitivity, and pressure in the case. Warm-up conditions are also determined.

Figure 5 presents a scheme of the experimental set-up. Oscillators under testing are put in a programmable climatic oven that enables to change automatically the environmental conditions. Frequency variations are compared with a reference oscillator. The data acquisition is made following a defined process by a computer.

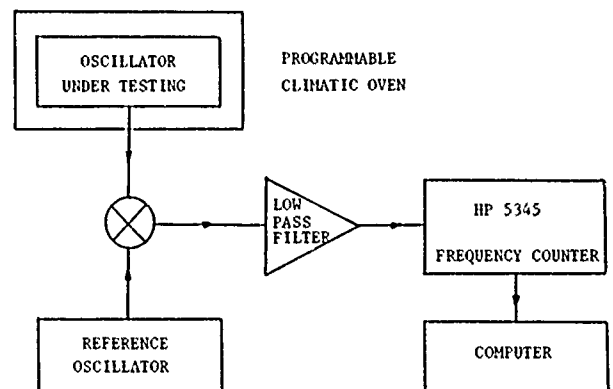


Fig. 5 - Scheme of the experimental set-up for oscillator tests

Short term stability :

Figure 6 presents the short term frequency stability of oscillators using a resonator CX-MV (6a), a resonator COMODOR (6b) and a ZT-cut resonator (6c). Frequency stability is always better than 1×10^{-9} in the range of 0.1 sec to 100 sec. The best short term stability is obtained with the ZT-cut resonator and corresponds to $2.5 \times 10^{-11}/10$ sec.

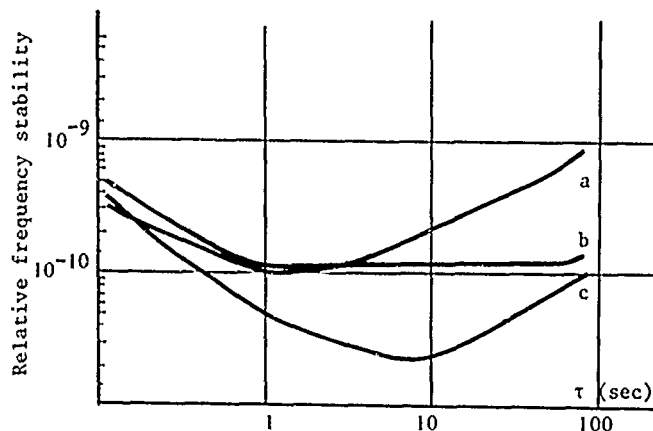


Fig. 6 - Short term frequency stability of the three different oscillators

Electrical power consumption versus temperature and power supply voltage :

The electrical consumption was measured on several oscillators at different temperatures in the range -40°C to $+55^\circ\text{C}$ when the supply voltage varies from 10 V to 20 V. Figure 7 shows that the total dissipated power is 500 mW at -40°C , 220 mW at $+20^\circ\text{C}$ and 80 mW at $+55^\circ\text{C}$.

The SARSAT program needs voltage up to 18 V.

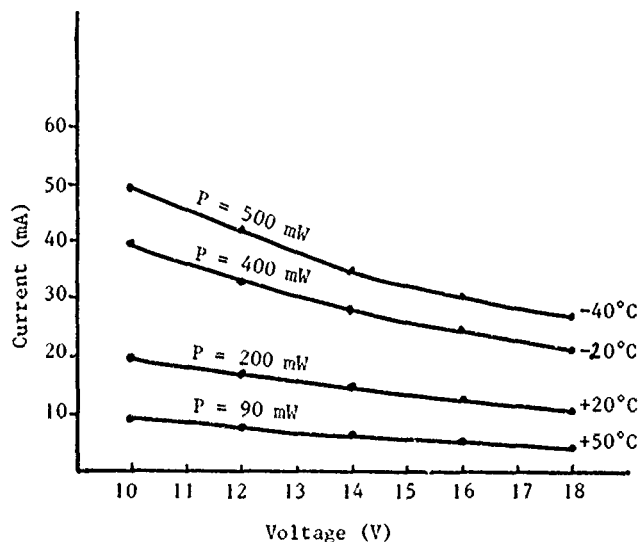


Fig. 7 - Electrical consumption versus temperature

Electrical power consumption versus gas pressure in the oscillator case

Figure 8 represents the dissipated power versus pressure within the case of the oscillator from atmospheric pressure 1×10^{-4} Torr. Measurements were performed at 20°C and the supply voltage was 12 V.

At low pressure reduces the consumption, as expected. But it must be noticed that it can be difficult to maintain a low pressure, inside the case, for long period of time, and therefore the oscillator is in practice operated at the atmospheric pressure.

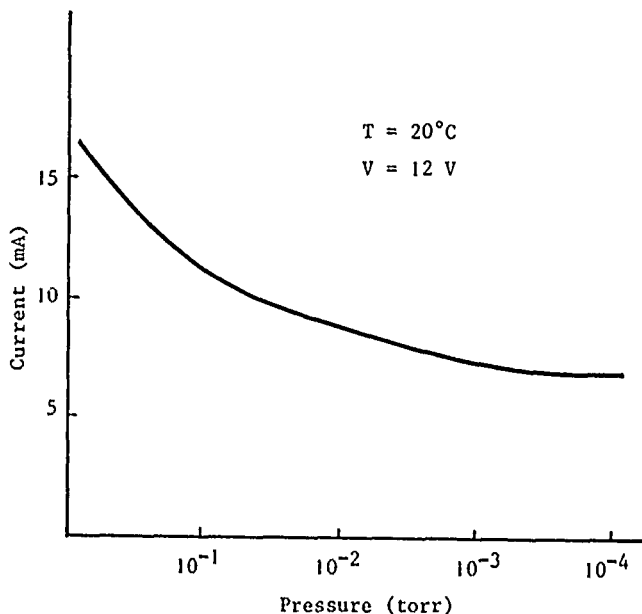


Fig. 8 - Electrical consumption versus gas pressure in the oscillator case

Oscillators warm-up time

The variations of the electrical consumption versus time when the oscillator is turned on, are represented on Figure 9 for different external temperature in the range -40°C to $+55^\circ\text{C}$. The power supply voltage is still equal to 12 V. The warm up behavior remains similar up to 20°C and shows an overshoot before the power stabilization is reached, after 120 s. Above 20°C the power decreases rapidly and the stabilization is obtained after 45 s.

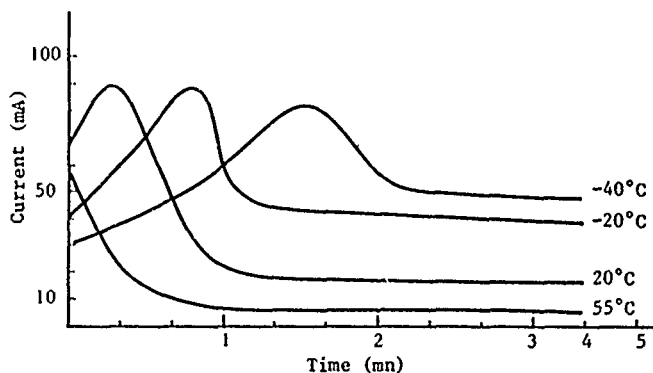


Fig. 9 - Electrical consumption as a function of warm-up time

Figure 10 represents the relative frequency shifts of the oscillator during the warm-up time : 1×10^{-8} is reached in 180 s and 1×10^{-9} in 10 mn. In this case, external temperature is equal to 20°C and the applied voltage is 12 V.

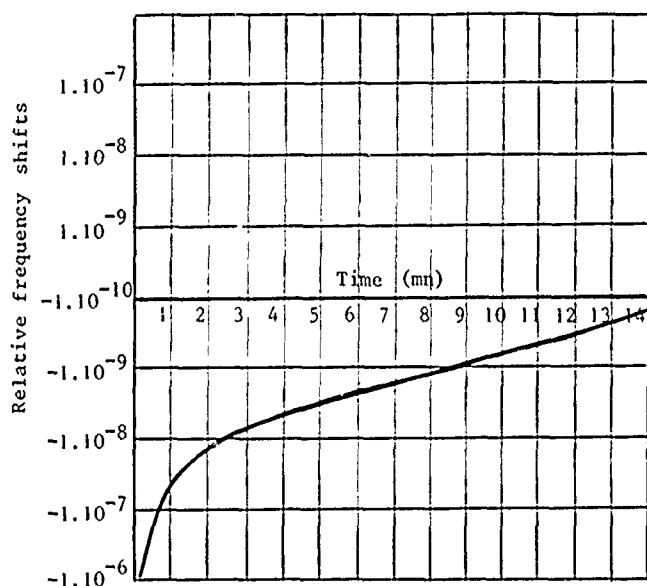


Fig. 10 - Warm-up frequency shifts

SARSAT oscillator

When the rescue buoy starts up, a signal at 406 MHz is emitted during 500 ms every 50 sec, and this is repeated 24 times ; then this cycle starts again until the buoy battery is empty. Each satellite makes a complete turn in 2 hours. As shown in Figure 11, when the buoy emits the signal, the battery voltage decreases by 2 volts.

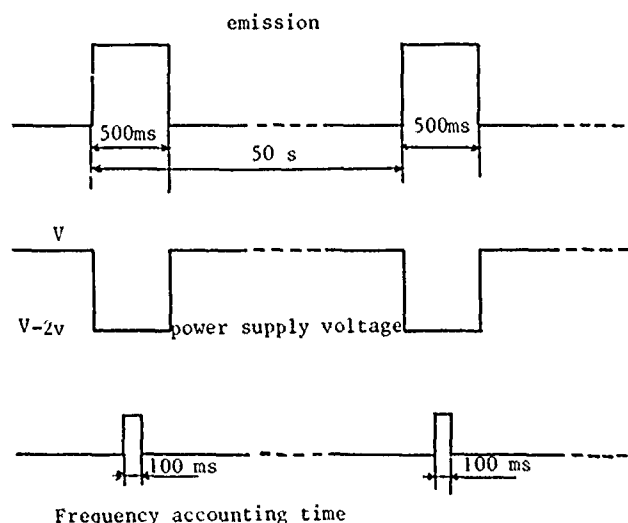


Fig. 11 - Timing diagram for SARSAT buoy

So, to characterize the oscillator in SARSAT conditions, 24 samples of the frequency are counted over 0.1 sec every 50 sec and a statistic is done with 48 groups of samples. Temperature is stabilized at -40°C during the first 20 mn, after that the measurements are performed with temperature varying from -40°C to 55°C with a slope of $0.1^\circ\text{C}/\text{mn}$.

Figure 12 represents the mean frequency of the oscillator measured in the previous conditions. SARSAT specifications correspond to 1×10^{-6} . The mean square deviations σ are plotted in Figure 13 and they must be smaller than 3×10^{-9} . Figure 14 represents the slope β of the mean line obtained by regression of each 24 frequency samples group, expressed in relative value per minute. It must be smaller than $\pm 1 \times 10^{-3}/\text{mn}$. Finally, Figure 15 shows the variations of dispersion σ_j of each 24 samples group around the linear regression line. σ_j limit is 3×10^{-9} .

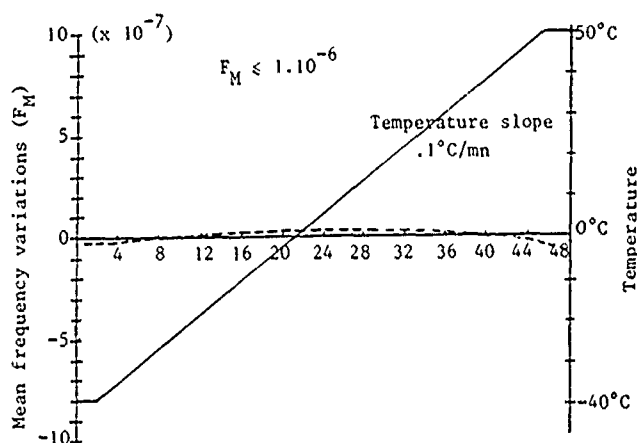


Fig. 12 - Mean frequency variations of SARSAT oscillator under temperature ramp

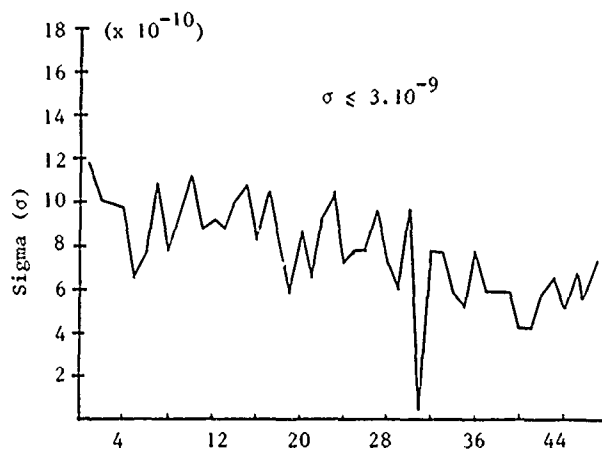


Fig. 13 - Mean square deviation σ o. SARSAT oscillator

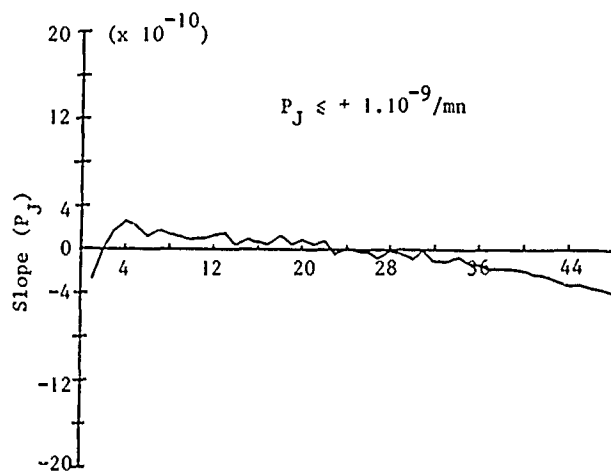


Fig. 14 - Slope P_J of the mean line of SARSAT oscillator

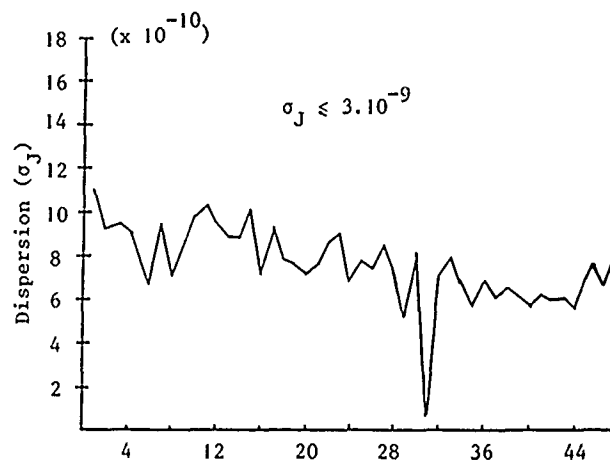


Fig. 15 - Dispersion variations σ_J of SARSAT oscillator

Table I summarized characteristics of the oscillators and their comparison with SARSAT requirements.

	SARSAT Requirements	Measured values
FREQUENCY	5075.3125 MHz $\pm 2.10^{-6}$	YES
SHORT TERM STABILITY	$< 2 \times 10^{-9} / 100 \text{ ms}$	$2 \times 10^{-10} / 100 \text{ ms}$
STABILITY WITH A 0.1 °C/mn TEMPERATURE GRADIENT	Slope $< 1 \times 10^{-9}/\text{mn}$ $\sigma < 3 \times 10^{-9}$	$\pm 5 \times 10^{-10}/\text{mn}$ 5×10^{-10}
AGING	$< \pm 1 \times 10^{-5} / 5 \text{ years}$	-
OUTPUT SIGNAL	TTL	TTL
ELECTRICAL CONSUMPTION	400 mW -40 °C 150 mW 25 °C	450 mW -40 °C 180 mW 25 °C
POWER SUPPLY VOLTAGE	10 V to 18 V (ΔV)	10 V to 18 V (ΔV)
WARM UP TIME	$< 10 \text{ mn at } -40 \text{ °C}$	$< 5 \text{ mn at } -40 \text{ °C}$

Table I

Acknowledgements

The authors wish to thank M. Fourcade of CNES-france for his help in the measurements in terms of the SARSAT specifications. This study was made under contract with the Centre National d'Etudes Spatiales and the Agence Nationale de Valorisation de la Recherche.

References

- ¹ J.J. Royer, "Rectangular AT-cut resonators", 27th AFCS, 30-34 (1973).*
- ² M. Onoe, M. Okazaki, "Miniature AT-cut strip resonators with tilted edges", 29th AFCS, 42-48 (1975).

* AFCS : Proceedings of the Annual Frequency Control Symposium.

- ³ A. Zumsteg, P. Suda, "Properties of a 4 MHz miniature flat rectangular quartz resonator vibrating in a coupled mode", 30th AFCS, 196-201 (1976).
- ⁴ M. Onoe, K. Kamada, M. Okazaki, F. Tajika, N. Manabe "4 MHz AT-cut strip resonator for wrist watch", 31st AFCS, 48-54 (1977).
- ⁵ R. Dinger, "A miniature quartz resonator vibrating at 1 MHz", 35th AFCS, 144-148 (1981).
- ⁶ W. Zingg, "Résonateurs et capteurs à quartz miniatures", 2nd Seminar on Frequency Standards, Characterization and Use. LPMO-Cetehor, Besançon, mars 1985.
- ⁷ J. Hermann, "A novel miniature ZI-cut resonator", 39th AFCS (1985).

A SHORT SURVEY OF FREQUENCY SYNTHESIZER TECHNIQUES

V. Reinhardt, K. Gould, K. McNab, and M. Bustamante
 Hughes Aircraft Company
 Space and Communications Group
 P.O. Box 92919
 Los Angeles, CA 90009

Introduction

Frequency synthesizers are the work horses of precise time and frequency. It is the job of frequency synthesizers to translate the performance of a reference oscillator to frequencies useful to the user. Frequency translators fall into two general classes: (fixed) reference generators and (adjustable) frequency synthesizers. In this paper, we will discuss the architectures and design techniques used in the construction of the second class, (adjustable) frequency synthesizers (though many of the techniques discussed are also used in the construction of reference generators). This paper is not meant to be an exhaustive treatment of frequency synthesizers, but only a brief and qualitative description of the basic frequency synthesizer architectures and techniques. Readers are referred to the bibliography at the end of the paper for more in depth material and for a quantitative analysis of the various architectures.

Frequency Synthesizer Design and Performance Parameters

Tables 1 and 2 summarize the important parameters which characterize the design and performance of frequency synthesizers. The tables are somewhat

TABLE 1. BASIC SYNTHESIZER PARAMETERS

Frequency range (maximum and minimum frequency)
Frequency resolution
Settling time
Phase/frequency stability (time domain)
Spectral purity (noise, harmonics, spurious sidebands)

TABLE 2. OTHER IMPORTANT SYNTHESIZER PARAMETERS

Of Output type (sine, square, pulse, other)
Output level
Amplitude performance (flatness, stability, accuracy)
Phase continuity (when frequency is changed)
Phase synchronization (to external epoch)
Operating environment (temperature, humidity, shock)
Environmental stability/performance
Size, weight, power consumption, cost, complexity
Required references (number and type)
Number of independent output frequencies
Manual or remote control (command format)

arbitrarily broken up into basic synthesizer parameters (Table 1) and other important parameters (Table 1). The basic synthesizer parameters in Table 1 are the obvious performance parameters that come to mind when defining the performance of a frequency synthesizer and need no further explanation.

The other important parameters listed in Table 2 are often equally as important as the basic parameters. Of course the type of output and the output levels required are important design parameters which often affect the complexity and type of design required. The amplitude performance of a frequency synthesizer, its flatness over the frequency range, its stability, and its accuracy are often as important as the basic phase and frequency performance parameters. The phase continuity, whether the signal must be continuous when the frequency is changed, and the phase synchronization requirements, whether the output must be synchronized to some reference epoch, are also of importance. The operating environment of the synthesizer is another important parameter. It is important in determining stresses the synthesizer must endure. The operating environment also affects the other parameters (environmental stability/performance).

Size, weight, power consumption, cost, and complexity constraints determine the type of designs that can be utilized. The number of required references strongly influences the overall complexity and size of the synthesizer because these references usually have to be generated with internal reference generators. The required number of independent output frequencies has great influence on the size penalty imposed by a complex set of reference generators; since one reference generator can drive many synthesizers, the size penalty imposed by complex reference generators becomes negligible as the number of required output frequencies grows. Last but not least, another important factor which affects circuit complexity is whether the output has to be controlled remotely or manually. For the remote case, the required command format also affects the complexity of the onboard command logic.

The above parameters together define the physical and performance constraints important in the design of frequency synthesizers. In the architectural description sections that follow, we will attempt to outline and compare the strengths and weakness of each architecture in terms of the above parameters along with the architectural descriptions. Ideally, one would like to quantitatively categorize the strengths and weaknesses of each architecture with a set of typical performance parameter values. We apologize for not presenting such charts in this paper, but we have not done so for two very good reasons. First, the state of the art of frequency synthesizers is so dynamic and the field is so diverse that such charts, even if they could be researched properly, would most likely be obsolete almost immediately. Second, and most important, the strengths and weakness of the various architectures are more properly defined in terms of tradeoffs between various competing parameters. One can, therefore, almost always increase the performance level of a particular parameter by sacrificing another, making a chart with fixed values impossible to create.

Frequency Synthesizer Techniques

In the architectural description sections that follow, we have separated the various frequency synthesizer techniques into three general classes: direct analog synthesis, indirect synthesis, and direct digital synthesis. In direct analog synthesis, the frequency of the reference oscillator or oscillators are translated directly, using analog techniques, without the use of an internal voltage controlled oscillator (VCO). In indirect synthesis, a VCO, which is phase or frequency locked to the reference oscillators, is the source of the synthesizer output frequency. In direct digital synthesis (DDS), the output waveform is directly synthesized digitally from the reference oscillator which is used as a clock for the digital operation. Because some DDS techniques are used with indirect synthesis, we will postpone our discussion of DDS techniques until after we have discussed indirect synthesis.

Direct Analog Synthesis

In direct analog synthesis, the frequency of the reference oscillator or oscillators are translated directly, using analog techniques. The basic direct analog synthesis techniques consist of frequency division, mixing, multiplication, switching, and filtering. The use of the word analog here is somewhat of a misnomer since frequency switching and division can be accomplished digitally. We use the word analog here to distinguish these techniques from direct digital synthesis. The distinction between direct analog and direct digital synthesis will become apparent after the section on direct digital synthesis.

Frequency Division

Frequency division is described by the basic operation:

$$f_o = f_a / N$$

where f_o is the output frequency, f_a is the input frequency and N is the division ratio.

Frequency division is accomplished by using a digital counter, a regenerative divider, or an injection locked oscillator. Before the advent of large scale integration (LSI), regenerative dividers and injection locked oscillators were popular because, for large division ratios, they are much less complex than digital dividers. Regenerative dividers and injection locked oscillators, however, are relatively narrowband devices and are susceptible to cycle skipping. With the advent of LSI, complex digital counters could be implemented monolithically in very small and low power packages. Digital dividers or counters are wideband devices, operating from dc to a maximum frequency determined by the type of digital logic used. (The current state of the art is about 10 GHz.) They have the disadvantage of outputting only a pulse or square wave, which has high harmonic content.

There are three types of digital counters used as frequency dividers: asynchronous or ripple counters, synchronous counters, and dual modulus counters. Ripple counters are relatively simple low power counters which are generally used as fixed frequency dividers. However, because the signal ripples through each of the stages sequentially, these counters can have relatively high levels of phase instabilities. In certain applications, these instabilities can be cleaned up by using a flip flop clocked by both the counter output and the input frequency. Presettable synchronous counters are used as variable frequency dividers. These counters tend to have lower levels of phase instabilities, but

higher levels of power consumption than ripple counters. A third type of counter used in variable division applications is called a dual modulus or swallow counter. This counter uses a high speed divide by $N/N+1$ counter which is controlled by lower speed presettable synchronous counters. It is used to produce very large variable division ratios with relatively low power consumption. (Lower than an equivalent high speed synchronous counter.) The dual modulus counter has the disadvantage of having a minimum division ratio as well as a maximum one, but this is not a problem in many applications.

All dividers, when used as frequency synthesizers, have the disadvantage of requiring a large N to achieve a fine frequency resolution. This means the reference frequency must be very large compared with the output frequency to achieve a high resolution.

Frequency Mixing

Frequency mixing is described by the basic operation:

$$f_o = f_a \pm f_b$$

where f_o is the output frequency and f_a and f_b are two input frequencies.

Frequency mixing is accomplished by using a non-linear element such as a diode or a transistor. The basic disadvantage of mixing is the presence of other mixing components of the form:

$$n f_a \pm m f_b$$

where n and m are integers. These generate spurious sidebands (spurs) in the desired signal. Frequency filtering can be used to eliminate these spurs, but at the price of narrowing the frequency range. A wideband solution to reducing unwanted spurs is to use single balanced, double balanced, triple balanced, and single sideband mixers. These mixers can reduce spurs to the -20 to -50 dBc level before filtering.

Frequency Multiplication

Frequency multiplication is described by the basic operation:

$$f_o = n f_a$$

where f_o is the output frequency, f_a is the input frequency, and n is an integer.

Direct analog frequency multiplication is accomplished using a non-linear element such as a diode, transistor, or varactor. The basic problem with frequency multiplication is the presence of spurs from the multiplication orders other than the desired one. These spurs can be on the order of or greater than the desired signal, and filtering is normally required to reduce them to acceptable levels. Balanced multipliers, such as full wave rectifiers, can be used to reduce the level of certain spurs before filtering. (For example, full wave rectifiers tend to suppress odd harmonic orders.)

Since a non-linear element produces both frequency mixing and multiplication, the two operations are sometimes combined in a single stage. This has the advantage of reducing circuit complexity, but usually produces higher spurs than the separated operations.

Frequency Switching

Frequency switching is the selection of one of sev-

eral input frequencies using switches. This operation is extremely straightforward, but requires as many references as output frequencies. The two basic parameters of importance in frequency switching are switch settling time and switch isolation in the open position. Switch settling time, of course, limits the synthesizer settling time. Switch isolation determines the level of spurs generated by the non-selected frequency sources leaking into the output. Three basic types of switches are used: mechanical, electro-mechanical, and electronic. Mechanical and electro-mechanical switches provide the highest degree of isolation, but are slow and bulky. Electronic switches can provide fast settling times and small size. However, in electronic switches, there tends to be a trade-off between size and power consumption and isolation. Monolithic FET switch arrays are extremely small and have very low power consumption, but have relatively low isolation. PIN and conventional diode switches have much higher levels of isolation, but are bulkier and have relatively high power consumption. Within the category of diode switches, there also tends to be a direct trade-off of isolation versus size and power consumption.

Frequency Filtering

Frequency filters are used to reduce the spurs and harmonics generated by the other techniques. Far away spurs and harmonics can be easily filtered using wide band filters without introducing many problems. Filtering nearby spurs, however, requires the use of narrow band filters. This can introduce the following problems:

1. The output frequency range is restricted to the pass band of the filter.
2. The settling time is inversely proportional to the pass band of the filter.
3. The phase stability, amplitude flatness, etc., can be affected by narrow pass bands.
4. Narrowband filters have large group delays which can cause instability in feedback paths.

There are two classes of filters used in frequency standards, passive and active. Passive filters used in synthesizers range from simple RC filters to complex SAW, crystal, and cavity filters. Filters require no power, but their size and weight tend to go up with the Q or required attenuation values. SAW filters have revolutionized the filter industry because of their small size to performance ratio compared with other types of filters.

Active filters of course require power. For lower frequencies, operational amplifier active filters are very useful in producing exotic filter responses. For higher frequencies, phase and frequency locked loops are used as active filters. These filters are covered in detail in the section on indirect synthesis. Active filters can also generate noise and spurs and can have oscillation problems due to the noise figure, distortion, and gain of the active elements.

Hybrid Direct Output Designs

The direct analog techniques listed above are very powerful when combined together in groups. Figure 1 shows a typical direct analog synthesizer which uses all the above techniques. This type of synthesizer is sometimes called a switch, divide, and mix synthesizer.

The main part of the synthesizer is the divide and mix section. This section consists of a sequence of

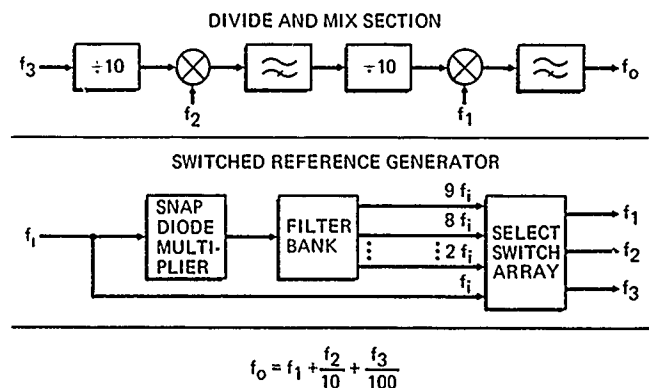


FIGURE 1. TYPICAL DIRECT ANALOG SYNTHESIZER

stages each containing a divider, a mixer, and a high pass filter. In each stage, an input frequency, such as f_3 , is divided by a number such as 10 in the divider, mixed with a second frequency, such as f_2 , in the mixer, and filtered to produce $f_3 + f_2/10$. This output frequency is input to the next divide, mix, and filter stage and so on as many times as desired to produce an output of the form:

$$f_0 = f_1 + f_2/10 + f_3/100 + \dots$$

The second part of the synthesizer, the selectable reference generator section, is used to generate f_1 , f_2 , f_3 , etc. This section consists of a snap diode multiplier, a SAW filter bank, and a select switch array. The snap diode multiplier and the SAW filter bank are used to generate a bank of 9 reference frequencies: $f_a, 2f_a, \dots, 9f_a$. This bank of reference frequencies is sent to the select switch array which routes any one of the 9 reference frequencies to the outputs f_1, f_2, f_3 , etc. Thus:

$$f_i = N_i f_a$$

where N_i is an integer. Using this, the output frequency becomes:

$$f_0 = (N_1 + N_2/10 + N_3/100 + \dots) f_a$$

Thus each reference frequency is used to select a digit in a decadic representation of the output frequency. In a practical design, additional switches are used to select zero possibilities for the digits, N_i , in the above formula.

Indirect Synthesis

Indirect synthesis utilizes an oscillator controlled by a phase lock loop (PLL) or a frequency lock loop (FLL) to generate its output frequency. The basic PLL/FLL is shown in Figure 2. It consists of the following:

1. A voltage controlled oscillator (VCO) which outputs a frequency f_0 .
2. One or more reference frequency inputs, f_a, f_b , etc.
3. A frequency translator defined by the function $T(f_0)$.
4. A phase discriminator (PD) or frequency discriminator (FD) which generates an output determined by the difference in phase or frequency between f_a and $T(f_0)$.
5. A loop filter which conditions the output of

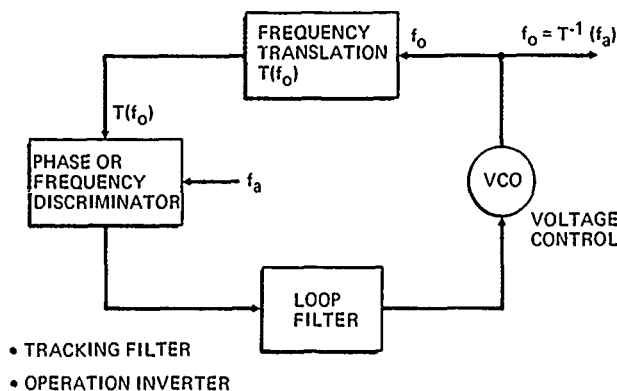


FIGURE 2. BASIC INDIRECT SYNTHESIZER

the phase or frequency discriminator before sending it to the voltage control input of the VCO

The FLL/PLL is a servo loop. When the loop locks, $T(f_o)$ becomes equal to f_a and the VCO output frequency is driven to:

$$f_o = T^{-1}(f_o)$$

where T^{-1} indicates the inverse function, not $1/T$.

The PLL/FLL has two very important properties for frequency synthesis, as an operation inverter and a tracking filter.

PLL/FLL as an Operation Inverter

The FLL/PLL inverts the frequency translation operation $T(f_a)$. That is, its output is $T^{-1}(f_a)$. Figure 3 shows a typical frequency translation loop synthesizer. Here f_b is subtracted from f_o and the result is divided by N before it is sent to FD or PD. Since the loop inverts this process, the resultant output is:

$$f_o = f_b + Nf_a$$

PLL/FLL as a Tracking Filter

The VCO's output tracks $T^{-1}(f_o)$ within a loop bandwidth B around the lock frequency. Outside this bandwidth, the VCO's performance is determined by the free running properties of the VCO. Thus the loop filters out any spurs, harmonics, and noise generated by the reference or the translation process $T(f_o)$ outside the loop bandwidth. This process creates a filter centered around the output frequency regardless of the value f_a or the translation function, $T(f)$, and thus effectively tracks with the output when f_a or $T(f)$ change. (Because $T(f)$ is inside the loop, care must be used when tracking changes in $T(f)$ since changing $T(f)$

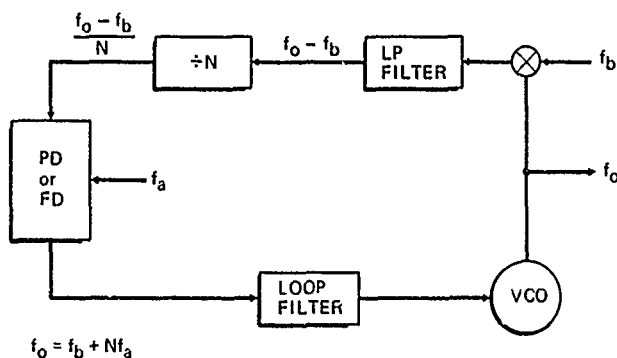


FIGURE 3. FREQUENCY TRANSLATION LOOP

also affects the loop response function as well as the center frequency.)

Effects of Loop Bandwidth on Synthesizer Performance

Setting the loop bandwidth frequency to the Fourier frequency (offset from the carrier) where the sum of the unfiltered phase noise spectral density from the reference and the loop electronics equals the free running VCO phase noise spectral density minimizes the total (integrated) phase noise. Also the loop bandwidth must be substantially lower than the reference frequency into the phase or frequency discriminator to keep harmonics of this reference frequency from generating spurious sidebands in the output frequency.

However, the settling time of a loop is inversely proportional to the loop bandwidth. Thus a settling time requirement sets a minimum loop bandwidth which can be used. This minimum loop bandwidth may be larger than the loop bandwidth required to minimize the phase noise, so the settling time requirement may force an increase in phase noise. This minimum loop bandwidth also constrains the lowest reference frequency which can be used at the phase or frequency discriminator without generating output spurs since this reference frequency must be substantially higher than the loop bandwidth. To minimize the constraints imposed by the settling time requirement, much effort has gone into reducing the loop settling time-to-bandwidth ratio. We will discuss this in more detail in a later section.

Phase vs Frequency Lock Loops

PLL's are much more popular in synthesizer applications today than FLL's for several reasons:

1. Good wideband PD's are relatively easy to make using digital techniques and balanced mixers, while FD's tends to be narrowband devices.

2. Loop instabilities and noise in an FLL are converted to frequency instabilities and noise, while the loop instabilities and noise in a PLL are only converted to phase instabilities and noise.

3. Tighter lock is maintained on the VCO by PLL's because a PLL locks the phase of the VCO to the phase of the reference input, while an FLL only locks the frequency of the VCO to the frequency of the reference input.

Since PLL's are by far the most popular of the two loops in frequency synthesizers, for the remainder of this section, we will limit our discussion to PLL's. Bear in mind, however, that many of the comments we will be making about PLL's also apply to FLL's.

Analog and Digital PLL's

PLL's in use today fall into two general groups: analog and digital loops. An analog PLL is shown in Figure 4 and a digital PLL is shown in Figure 5. In the analog PLL shown in Figure 4, the phase detector or phase discriminator outputs an analog voltage which depends on the phase difference between the reference and the feedback VCO signal. This analog voltage is processed in an analog loop filter and then used to control the VCO.

In the digital PLL shown in Figure 5, a phase error quantizer, outputs a digital word which depends on the difference in phase between the reference and the feedback VCO signal. This quantized phase error word is sent to a hardware or software digital loop filter which processes the error word and outputs a VCO control word to a digital-to-analog converter (DAC).

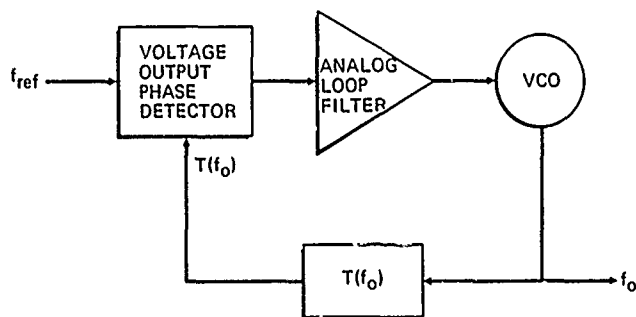


FIGURE 4. ANALOG PHASE LOCK LOOP

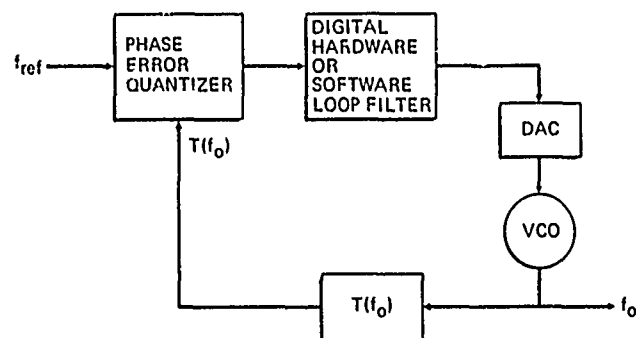


FIGURE 5. DIGITAL PHASE LOCK LOOP

The DAC then turns the control word into a control voltage for the VCO.

We will compare the pros and cons of analog versus digital loops in a later section, but before we do, it is important to explain about the techniques used to improve the settling time of PLL's.

Fast Loop Settling Techniques

As mentioned previously, because it is important to minimize the constraints imposed by settling time on loop bandwidth, much design effort has gone into developing techniques which reduce the settling time of a loop for a given loop bandwidth. Figure 6 outlines some of these techniques as used in an analog loop. A short discussion of each of the techniques, as applied in an analog PLL follows. Each technique has an equivalent form in a digital PLL.

Pretune DAC. The settling time of a loop is directly related to initial frequency offset of the VCO when the frequency of the loop is changed. Since the VCO frequency output versus control voltage input curve is known to some level, the initial frequency offset

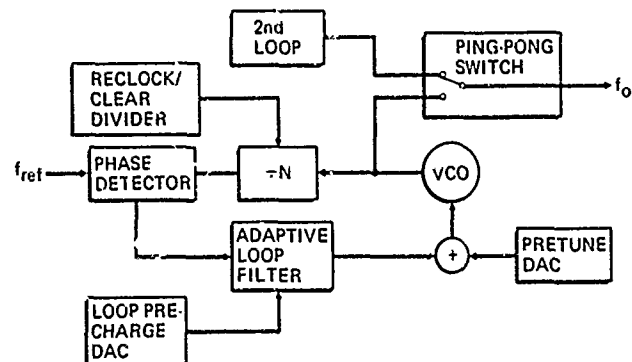


FIGURE 6. FAST LOOP SETTLING TECHNIQUES

can be reduced drastically by using a digital to analog converter (DAC) to pretune the VCO to approximately the right frequency when the loop frequency is changed.

Loop Precharge DAC. A related technique is the loop precharge DAC. In a second order loop, the voltage output of the loop filter, which generates the control voltage for the VCO, is determined by the charge on a capacitor in the loop filter. In order to reduce the initial frequency offset of the VCO when the loop frequency is changed, one can use a DAC to precharge this loop capacitor. The advantage of using the precharge DAC over the pretune DAC is that the precharge DAC is disconnected from the system once it has charged the loop capacitor. This means that the DAC voltage noise does not increase the synthesizer steady state phase noise in the case of the precharge DAC. The principal disadvantage of the precharge DAC is that it is much more complicated to implement than a pretune DAC.

Reclock/Clear Divider. Reclocking and clearing the divider is a technique which improves settling time by reducing the initial phase error in the loop when the loop frequency is changed. When the loop is locked in the steady state condition, the divide by N counter clears and outputs a pulse to the phase detector at the same instant that the reference frequency supplies a pulse to the phase detector. This is another way of stating that there is zero phase error between the reference frequency and the divided down VCO frequency. Thus the size of the divider count at the reference clock epoch is a measure of the phase error in the loop at that instant. In the initial stages of changing the frequency of the loop, the divide by N counter usually accumulates a large random count due to the fact that the VCO frequency is varying widely. This random count represents an initial phase error which the loop must track out before it settles. Also since N is usually quite large, this initial phase error can be equivalent to many cycles of phase error at the VCO frequency, which can take a long time to track out. One way of minimizing this initial phase error is to relock and clear the divider with the reference frequency some time after the initial stages of loop settling. This greatly reduces the initial phase error of the loop, and thus reduces the loop settling time significantly.

Adaptive Loop. A wideband loop settles faster than a narrowband loop. If one were to use a wideband loop in the early stages of loop settling and somehow switch to a narrowband loop as the loop settles, one could speed up the settling process. This is called an adaptive loop. One way the switchover from wideband to narrowband filtering is accomplished is by putting diodes in the loop filter which change their impedance with loop error.

Ping-Pong Loop. One can completely side step the settling time problem by putting two loops in a synthesizer and switching between them. This is called a ping-pong loop. In a typical ping-pong sequence, the ping-pong switch changes over to a new loop after it has had time to settle. The old loop then is immediately changed to a new frequency and the ping-pong switch is again switched after the loop settles. In a ping-pong loop, the loop settling time now limits the switching period rather than the settling time itself.

Pros and Cons of Digital and Analog Loops

Now let us proceed with a discussion of the pros and cons of analog versus digital loops. One disadvantage of analog loops is the fact that $T(f_o)$ must be equal to the reference frequency for the loop to lock. This is because a dc voltage must be output out of the

phase detector for an analog loop to be in a steady state locked condition. Conversely, in a digital loop, a changing phase error word out of the phase quantizer can be corrected for by subtracting a changing word in the digital loop filter. This means a digital loop can lock with $T(f_o)$ not equal to the reference frequency. Because of this, a digital loop can:

1. Track multiple independent reference frequencies in a single loop just by adding extra phase error quantizers
2. Perform part of the frequency translation in the digital processor of the loop filter

Another disadvantage of analog loops is that they are susceptible to burst error. Burst error is a large noise spike or other effect which causes the loop capacitor to discharge. The problem with this is two fold. First, this causes the VCO frequency to go to its maximum or minimum frequency. Second, it now takes the full settling time of the loop to recover to its original state. In a digital loop, this susceptibility to burst error can be made virtually nil by using limit algorithms which limit the rate of change of the equivalent loop capacitor and by using special memory to store the equivalent of the capacitor charge state.

A third disadvantage of analog loops is that fast loop settling techniques are very hardware intensive. In digital loops, many of these techniques can be implemented very easily with software algorithms or digital hardware.

One big advantage of digital loops is that processing time limits the maximum loop bandwidth. This problem is more severe in software loops, which have lower operating speeds than hardware loops.

A final disadvantage of digital loops occurs because the control DAC quantizes the voltage used to control the VCO frequency. This means there is a minimum frequency step size that the loop can implement. When this minimum step size is larger than the RMS frequency jitter of the loop, the digital loop can undergo limit cycle oscillations which generate large spurious sidebands. This limit cycle problem, however, can be eliminated by dithering the loop with a high frequency error signal, causing the DAC to jump between two adjacent states at the high frequency. This high frequency jitter can then be filtered out before the DAC voltage goes to the VCO by adding an analog low pass filter in between the DAC and the VCO, so the digital loop behaves like a switching regulator.

Direct Digital Synthesis

Direct digital synthesizer (DDS) designs in the technical literature fall into six major categories: pulse output DDS's, fractional divider or pulse snatching DDS's, sine output DDS's, triangle output DDS's, phase interpolation DDS's, and jitter injection DDS's. A brief discussion of each category follows.

Pulse Output DDS

The pulse output DDS (Kodanov, 1981; Peters, 1982) is the simplest of the five DDS categories. As shown in Figure 7, it merely consists of an N bit accumulator set up to add a frequency word, K, in an accumulator once every clock period, T_c . That is, if the current accumulator register value is R, once every T_c , the accumulator performs the operation:

$$R + K + R$$

in modulo 2^N arithmetic. Note that for this addition

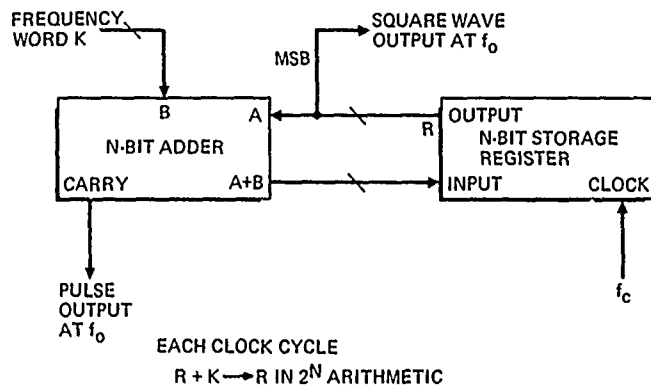


FIGURE 7. PULSE OUTPUT DIRECT DIGITAL SYNTHESIZER

process, the accumulator will overflow, on average, once every $2^N/K$ clock periods, so the average frequency of overflows will be:

$$f_o = F f_c$$

where f_c , the clock frequency, is $1/T_c$, and where the fractional output frequency, F, is given by:

$$F = K/2^N$$

The frequency output of this synthesizer is merely the carry output of the accumulator for a pulse output or the most significant bit (MSB) of the accumulator for an approximate square wave output. A typical example of the output of a pulse output DDS is shown in Figure 8.

The basic problem with this architecture is that it has very high levels of spurs and phase jitter.

Fractional Divider or Pulse Swallowing DDS

The fractional divider (Hassun, 1984; Nazarenko, 1982; Nissonevitch, 1978; No Author, 1982; Schineller, 1982; Rohde, 1981; Rohde, 1983) or pulse swallower

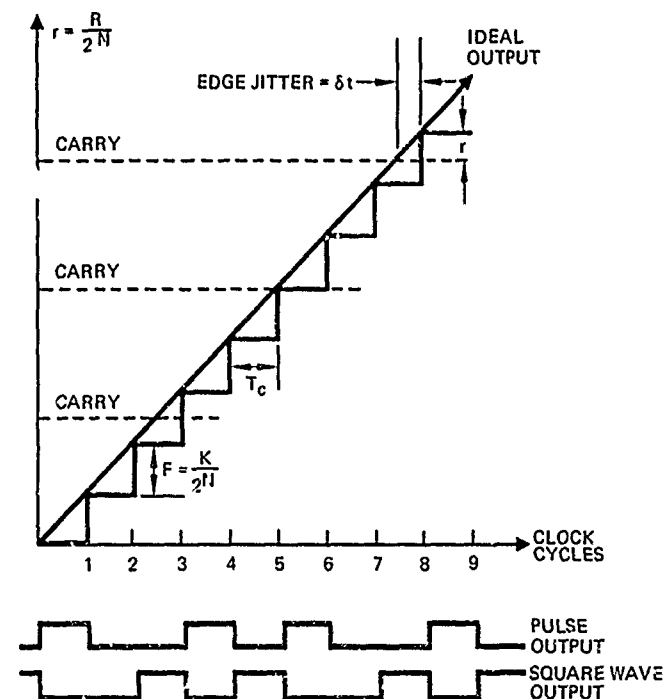


FIGURE 8. TYPICAL PULSE OUTPUT DIRECT DIGITAL SYNTHESIS WAVEFORM

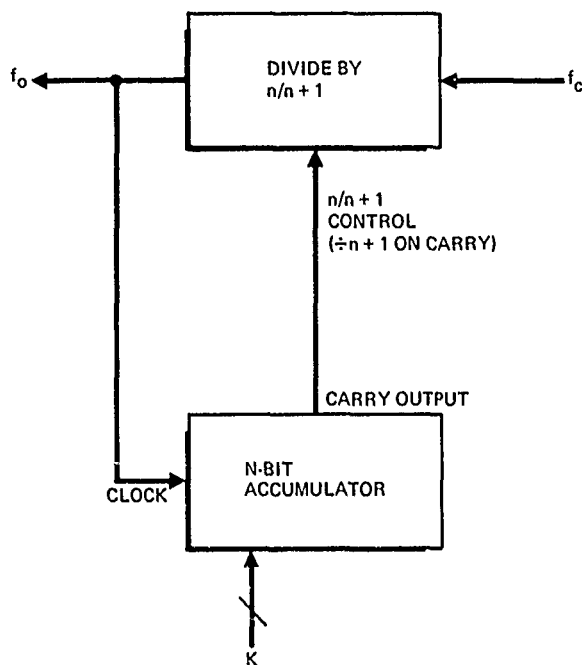


FIGURE 9. FRACTIONAL DIVIDER OR PULSE SWALLOWING DIRECT DIGITAL SYNTHESIZER

(Kohler, 1983) is a variation on the pulse output DDS. A block diagram of the fractional divider DDS is shown in Figure 9. In this type of DDS, the accumulator carry output is used to drive the $n/n+1$ control line of a divide-by- $n/n+1$ counter so that $n+1$ division occurs on a carry. The accumulator, in this case, is clocked by the output of the divider, f_o . The divider is clocked by the f_c input, and the output of the DDS is f_o . One can show that on average, the output frequency is:

$$f_o = f_c / (n+F)$$

(Notice that, in this case, F determines the fractional part of the division.) A typical example of the output of a fractional divider DDS is shown in Figure 10.

This type of DDS also has high levels of spurs and phase jitter.

Sine Output DDS

The sine output DDS produces a smoother, more sine-like signal by adding a sine look-up table and a digital to analog converter (DAC) to the pulse type DDS (Tierney, 1971; Gorski-Popiel, 1975; Rabiner, 1975; Galbraith, 1982; Hoppes, 1982; Kaiser, 1985; Crowley,

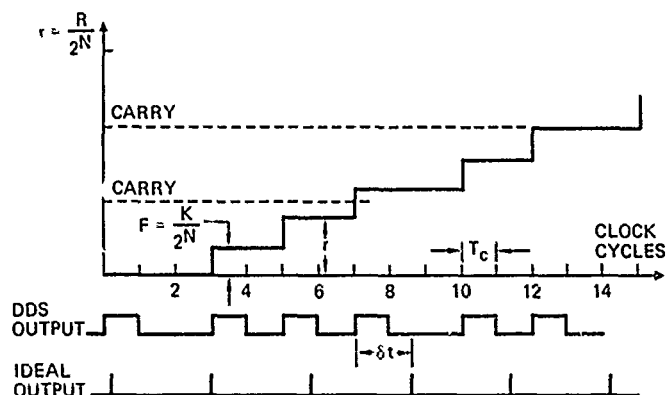


FIGURE 10. FRACTIONAL DIVIDER TYPICAL OUTPUT WAVEFORM

1982; Rohde, 1983). A block diagram of the sine output DDS is shown in Figure 11. The sine look-up table computes $\sin(2\pi R/2N)$ to the resolution of the sine table. The output of the sine-table is then sent to a DAC which outputs a voltage proportional to the sine table value to the M -bit resolution of the DAC. The result of this process is to produce a stepped sine wave output which has very low levels of spurs and phase jitter when low pass filtered. Figure 12 shows a typical stepped output of a sine output DDS.

The levels of spurs and phase noise in this DDS are directly related to the accuracy and resolution of the sine table and DAC. Generating a high resolution sine value directly from a single table usually requires a prohibitively large ROM, so techniques have been developed to reduce the ROM requirements by computing the sine value from several lower resolution tables (Sunderland 1984).

Phase Interpolation DDS

A phase interpolation DDS (Hassun, 1984; Kochemasov, 1982; DesBrisay, 1970; Crowley, 1982; Schineller, 1982; Rohde, 1981; Rohde, 1983; Gillette, 1969; Nossen, 1980; Bjerde, 1976a; Bjerde, 1976b) is similar to the sine output DDS in that it produces

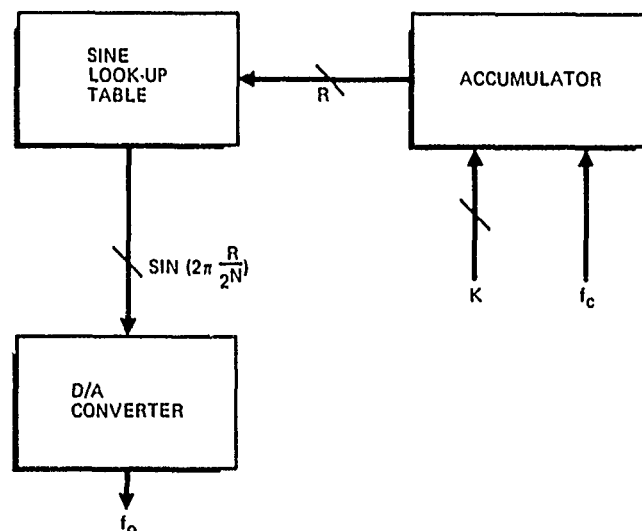


FIGURE 11. SINE OUTPUT DIRECT DIGITAL SYNTHESIZER

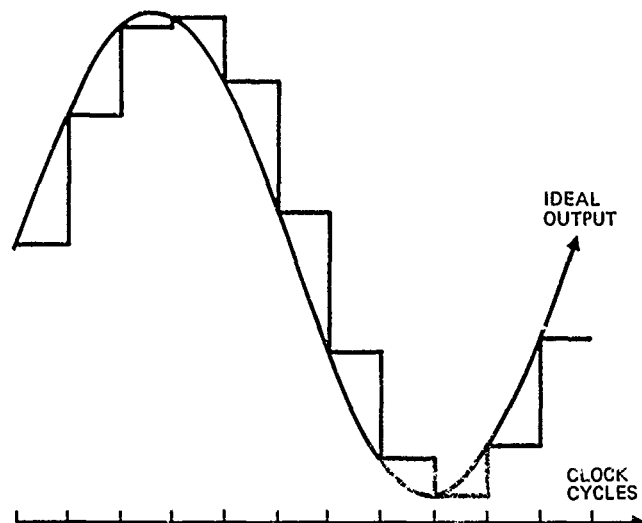


FIGURE 12. TYPICAL SINE OUTPUT DIRECT DIGITAL SYNTHESIS WAVEFORM

lower spurs, but it does not require a sine look-up table. Two versions are shown in Figures 13 and 14. The phase interpolation DDS utilizes the fact that, whenever an output transition occurs in a pulse output DDS or a fractional divider, the accumulator register value R is proportional to the time or phase difference between the output transitions of the DDS and that of an ideal frequency generator. Thus if R is used to phase shift or delay the output of a pulse output or fractional divider DDS, lower phase is jitter and spurs will result. In Figure 13, the output phase shifted using a phase lock loop (PLL) consisting of a linear phase detector, a differential loop amplifier, and a DAC driven by the DDS accumulator register (Hassun, 1984; Gillette, 1969; Nossen, 1980; Rohde, 1981; Rohde, 1983; Bjerrede, 1976a; Bjerrede, 1976b; Schineller, 1982; Crowley, 1982). In Figure 14 either a digitally controlled phase shifter (DesBrisay, 1970) or a digitally controlled delay generator (Kochemasov, 1982) driven by the DDS accumulator register are used to directly phase shift or delay the output. Figure 15 shows a typical output wave form from a digital phase shifter type of phase interpolation DDS. The phase jitter and spur level reductions that are achievable with phase interpolation DDS's are limited by the linearity, accuracy, and resolution of the digital-to-phase or delay conversion process.

A simplified version of the phase interpolation DDS with much narrower frequency range is the phase microstepper (Lavanceau, 1985; DesBrisay, 1970). This device uses a digitally controlled phase shifter operating off a single clock frequency produce small variations in the clock frequency with extremely high resolution. A typical phase microstepper produces 5 MHz plus or minus one part in $1E-7$ with a fractional frequency resolution of $1E-17$ (Lavanceau, 1985).

Triangle Output DDS

A triangle output DDS is another variation of a

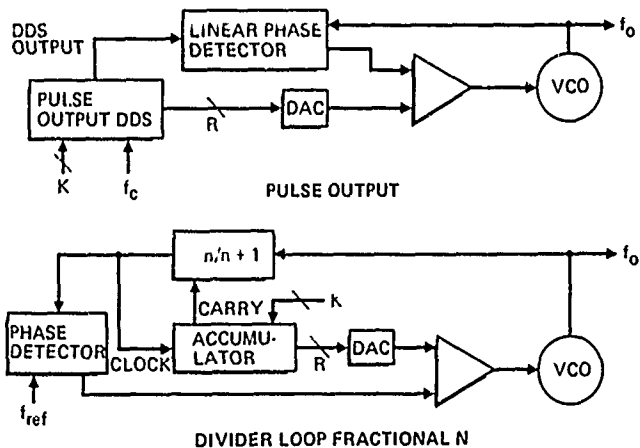


FIGURE 13. PHASE LOCK LOOP TYPE OF PHASE INTERPOLATION DIRECT DIGITAL SYNTHESIZER

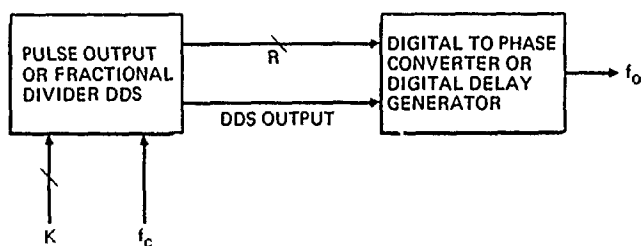


FIGURE 14. DIRECT OUTPUT TYPE OF PHASE INTERPOLATION DIRECT DIGITAL SYNTHESIZER

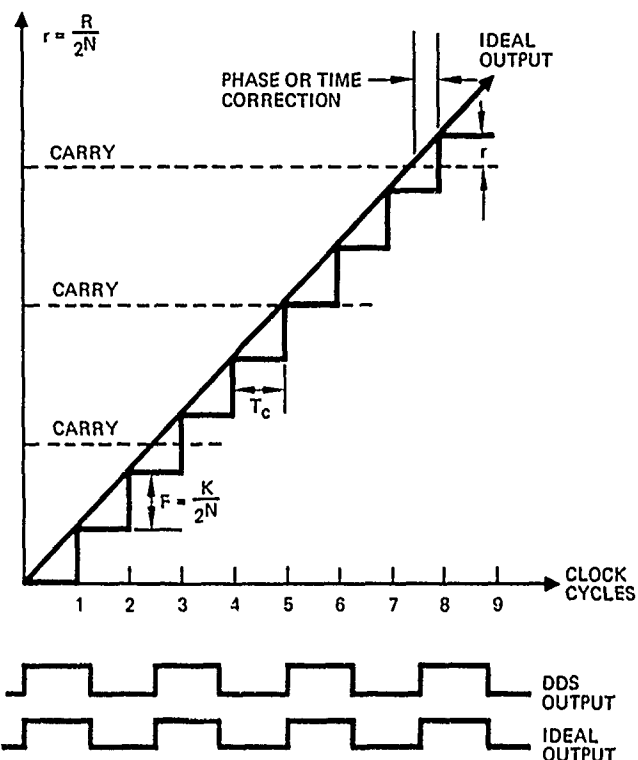


FIGURE 15. TYPICAL PHASE INTERPOLATION DIRECT DIGITAL SYNTHESIS OUTPUT

sine output DDS which does not require a sine table (DesBrisay, 1984). Its block diagram is shown in Figure 16 along with a typical output. In this type of DDS, the accumulator register value R of a pulse output DDS is used to drive a DAC directly after passing through a bit complement logic circuit. This produces a stepped triangle wave output. This triangle wave output has lower spurs than the outputs of a fractional divider or pulse output DDS.

Jitter Injection DDS

Wheatley has patented (Wheatley, 1983) a random jitter injection technique for use on a pulse output DDS which reduces the size of the spectral spurs in the output. This technique reduces the spurs by destroying the periodicity of the phase deviation patterns of the output transitions (Wheatley, 1981). The technique has two embodiments as shown in Figures 17 and 18. The periodicity in both embodiments is destroyed by randomizing the accumulation process with a digital random word. In doing so, one trades off spur levels for

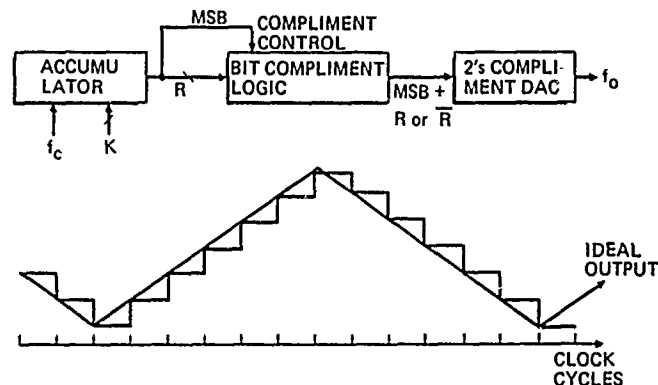


FIGURE 16. TRIANGLE OUTPUT DIRECT DIGITAL SYNTHESIS AND TYPICAL OUTPUT

increased random phase noise. Figure 19 shows a sample spectrum of a pulse output DDS with and without the Wheatley technique.

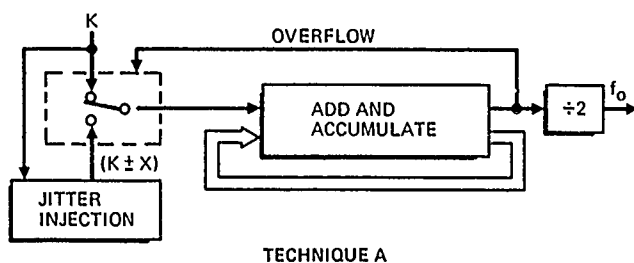


FIGURE 17. WHEATLEY RANDOM JITTERING TECHNIQUE A

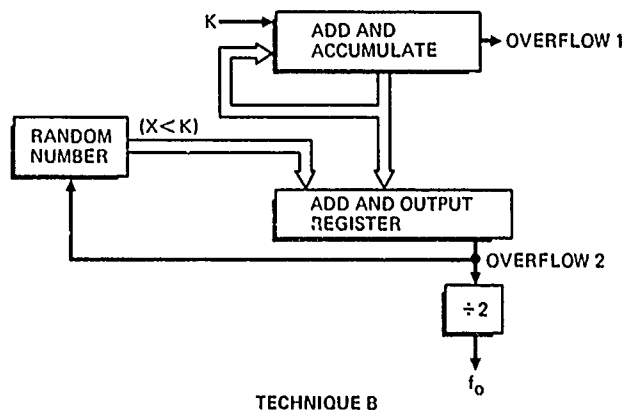


FIGURE 18. WHEATLEY RANDOM JITTERING TECHNIQUE B

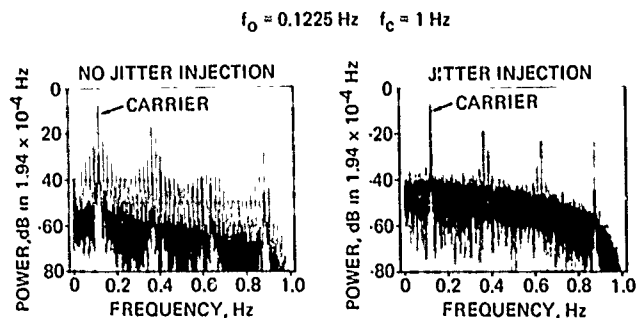


FIGURE 19. SIMULATION OF TYPICAL JITTER INJECTION DIRECT DIGITAL SYNTHESIS OUTPUT*

*COURTESY OF J. DAMIR, A. STRODTBECK, AND M. FASHANO, HUGHES AIRCRAFT COMPANY

References

- General Frequency Synthesis**
- J. Gorsky-Popiel, Ed., *Frequency Synthesis and Applications*, IEEE Press, 1975.
- V. F. Kroupa, *Frequency Synthesis*, Wiley, 1973.
- V. F. Kroupa, "Low-Noise Microwave Synthesizer Design Principles," *IEEE Proc.*, Volume 130, Part H, No. 7, p. 483, October 1983.
- V. Manassewitch, *Frequency Synthesizers Theory and Design*, Wiley, 1980.
- J. Noordanus, "Frequency Synthesizers - A Survey of Techniques," *IEEE Trans. Comm. Tech.*, Com-17, No. 2, April 1969.
- K. A. Samoylo, et al, "Frequency Synthesizers," *Radiotekhnika (USSR)*, Volume 30, No. 5, 1975.
- K. R. Slinn, et al, "Low-Noise Synthesizers for Radar and Communications," *IEEE Proc. (Great Britain)*, Volume 130, Part H, No. 7, December 1983.
- "Technology Closeup: Synthesizers," *Microwaves and RF*, June 1984.
- Direct Analog Synthesis**
- J. Schiffer, "EW Systems Synthesizers Find a New Design Approach," *MSN*, December 1981.
- Indirect Synthesis and Phase Lock Loops**
- J. M. Bryant, "Digital Frequency Synthesizer Design," *Wireless World*, June 1982.
- P. Connell and B. Hull, "Synthesized Source Joins Analog Purity, Digital Speed," *Electronic Design*, 30 September 1982.
- J.A. Crawford, "Understanding the Specifics of Sampling in Synthesis," *Microwaves and RF*, August 1984.
- J.A. Crawford, "Extending Sampling to Type II Phase-Locked Loops," *Microwaves and RF*, September 1984.
- W. F. Egan, *Frequency Synthesis by Phase Lock*, Wiley, 1981.
- F. M. Gardner, *Phaselock Techniques*, Wiley, 1979.
- F. M. Gardner, "Charge-Pump Phase-Lock Loops," *IEEE Trans. on Comm.*, Com-28, No. 11, November 1980.
- S. C. Gupta, "Phase-Locked Loops," *Proc. IEEE*, Volume 63, No. 2, February 1975.
- B. S. Glance, "New Phase-Lock Loop Circuit Providing Very Fast Acquisition Time," *IEEE Trans. MTT*, MTT-33, No. 9, September 1985.
- V. F. Kroupa, "Noise Properties of PLL Systems," *IEEE Trans. on Comm.*, Volume COM-30, No. 10, p. 2244, October 1982.
- U. Rohde, *Digital PLL Frequency Synthesizers*, Prentice-Hall, 1983.
- M. Sarhadi, "A Fast Frequency Synthesizer," *IEEE Colloq. on Phase-Locked Techniques*, London, England, 26 March 1980.
- M. J. Underhill, "Wide Range Frequency Synthesizers with Improved Dynamic Performance," *The Radio and Electronic Engineer (Great Britain)*, Volume 50, No. 6, June 1980.
- Digital Loops and Nonlinear Loops**
- J. Cessna, "Phase Noise and Transient Times for a Binary Quantized Digital Phase-Locked Loop in White Gaussian Noise," *IEEE Trans. on Comm.*, Com-20, No. 2, April 1972.
- N. A. D'Andrea and F. Russo, "A Binary Quantized Digital Phase Locked Loop," *IEEE Trans. on Comm.*, Com-26, No. 9, September 1978.
- N. A. D'Andrea and F. Russo, "Multilevel Quantized DPLL Behavior with Phase- and Frequency-Step Plus Noise Input," *IEEE Trans. on Comm.*, Com-28, No. 8, August 1980.

D. P. Garg, "Developments in Nonlinear Controller Synthesis: An Overview," *J. Dyn. Syst. Meas. and Cont.*, March 1978.

J. Garodnick, et al, "Response of an All Digital Phase-Locked Loop," *IEEE Trans. on Comm.*, Com-22, No. 6, June 1974.

G. S. Gill and S. C. Gupta, "First-Order Discrete Phase-Locked Loop with Applications to Demodulation of Angle-Modulated Carrier," *IEEE Trans. on Comm.*, June 1972.

G. S. Gill and S. C. Gupta, "On Higher Order Discrete Phase-Locked Loops," *IEEE Trans. on Aerospace and El. Systems*, AES-8, No. 5, September 1972.

J. Holmes, "Performance of a First-Order Transition Sampling Digital Phase-Locked Loop Using Random-Walk Models," *IEEE Trans. on Comm.*, Com-20, No. 2, April 1972.

G. T. Hurst and S. C. Gupta, "Quantizing and Sampling Considerations in Digital Phase-Locked Loops," *IEEE Trans. on Comm.*, January 1974.

W. C. Lindsay and C. M. Chie, "A Survey of Digital Phase-Locked Loops," *IEEE Proc.*, Volume 69, No. 4, April 1981.

G. Pasternack and R.L. Whalin, "Analysis and Synthesis of a Digital Phase-Locked Loop for FM Demodulation," *The Bell System Journal*, December 1968.

C.P. Reddy and S.C. Gupta, "A Class of All Digital Phase Locked Loops: Modelling and Analysis," *IEEE Trans. IECI*, IECI-20, No. 4, November 1973.

R.J. Simpson and H.M. Power, "Applications of High Frequency Signal Injection in Non-Linear Systems," *Int. J. Control*, Volume 26, No. 6, 1977.

A. Weinberg and B. Liu, "Discrete Time Analysis of Nonuniform Sampling First and Second Order Digital Phase Lock Loops," *IEEE Trans. on Comm.*, Com-22, No. 2, February 1974.

Frequency Lock Loops

Y. Bas-Ness and H. Bunin, "Interference in Frequency-Locked Doppler Tracking Loops," *IEEE Trans. Aerospace and El. Sys.*, AES-20, No. 5, September 1984.

A.D. Delagrange, "Lock Onto Frequency," *ED*, Volume 13, 21 June 1977.

Z. Galini and R.C. Waterman, "Automatic Frequency Acquisition Demonstrated Using Step Function Frequency Discriminator," *MSN*, November 1983.

C.W. Malinowski and H. Rinderle, "Designing Tracking Error-Free Varactor Tuners with Frequency Locked Loop Integrated Circuits," *IEEE Trans. Consumer El.*, CE-26, February 1980.

C.W. Malinowski, et al, "A Novel Frequency-Processing Method and its Implications on Future Tuning Systems," *IEEE Trans. Consumer El.*, CE-25, August 1979.

L.K. Regenbogen, "A Loose-Locked Oscillator," *The Radio and Electronics Engineer* (Great Britain), Volume 48, No. 3, March 1978.

T. Yuge, et al, "Phase-Following Digital Frequency-Locked Loop System," *Electronics and Communications in Japan*, Volume J59-A, No. 4, 1976.

F. Sutaria and D. Vazirani, "Modern Techniques of FLL Tuning," *IEEE Trans. Cons. El.*, CE-28, No. 4, November 1982.

Direct Digital Synthesizers

B. Bjerede and G. Fisher, "A New Phase Accumulator Approach to Frequency Synthesis," *Proc. IEEE NAECON 76*, p. 928, May 1976.

B. Bjerede and G. Fisher, "An Efficient Hardware Implementation for High Resolution Frequency Synthesis," *Proc. Frequency Control Symposium (USERADCOM)*, Atlantic City, June 1976.

C. Cole, "Design of a Direct Digital Synthesizer," Masters Thesis, Massachusetts Institute of Technology, 1982.

A.T. Crowley, et al, "Frequency Synthesizer," RCA Advanced Technology Labs, Camden, New Jersey, 1st Interim Report to ERADCOM, Ft. Monmouth (DELET-TR81-0376-1), February 1982.

G. DesBrisay Jr., et al, "Signal Converter - Low PRF 481083-151 Unit Design Description," Hughes Technical Memorandum 70-08, September 1970.

G. DesBrisay Jr., Hughes Aircraft Company, private communication, 1984. Hughes Aircraft Space and Communications Group has flown triangle output DDSs.

G.C. Gillette, "Digiphase Principle," *Frequency Technology*, August 1969.

J. Galbraith, "Generate Sine Wave by Direct Table Look-up," *EDN*, p. 101, 28 April 1982.

R. Hassun, "The Common Denominators in Fractional N," *Microwaves and RF*, p. 107, June 1984.

P.E.C. Hoppes, "A Radiation Hardened Low Power Numerically-Controlled Oscillator," *Proc. IEEE 1982 Custom Integrated Circuits Conference*, Rochester, New York, 17-19 May 1982.

H.W. Kaiser, et al, "SOS Frequency Synthesizer," RCA Advanced Technology Labs, Moorestown, New Jersey, Final Report to ERADCOM, Fort Monmouth (DELET-TR81-0376-F), January 1985.

V.N. Kochemasov and A.N. Fadeev, "Digital Computer Synthesizers of Two Level Signals with Phase Error Compensation," *Radiotekhnika* (USSR), p. 15, October 1982.

A. Kohler and R. Schiffl, "Basic Electronic Circuits: The Synthesizer II," *Funkschau* (Germany), No. 8, p. 73, 15 April 1983.

V.P. Kodanov and Y.D. Kononov, "Passive Digital Frequency Synthesizer Based on Standard Digital Integrator," *Elektrosvyaz* (USSR), Volume 35, No. 6, June 1981 and *Trans. in Telecom. and Radio Eng. (USA)*, Part 1, Volume 35, No. 6, June 1981.

J. Lavanceau, Private Communication, Time System Technology, Lake Grove, New York, Bethesda, Maryland, 1985. Time Systems Technology produces the model 6490 phase microstepper.

Nazarenko, et al, "A High Speed Digital Frequency Synthesizer with High Resolution," *Radiotekhnika* (USSR), Volume 37, No. 6, June 1982.

D.G. Nissonevitch, et al, "Classification of Digital Circuits for Frequency Synthesis," *Elektrosvyaz*

(USSR), Volume 32, No. 3, 1978 and *Telecom. and Radio Eng.*, Part 1, Volume 32, No. 3, 1978.

Author not cited, "Single Chip Frequency Synthesizer Based on XR2240," *Elektron. Appl.* (Germany), Volume 14, No. 7, July 1982.

E.J. Nossen, "Digital Frequency Synthesis," United States Patent 4,206,425, 3 June 1980.

H. Peters, Private Communication, Sigma Tau Standards Corporation, Tuscaloosa, Alabama, 1982. Sigma Tau Standards has used pulse output DDSs in their hydrogen maser frequency standards.

L. Rabiner and B. Gold, *Theory and Applications of Digital Signal Processing*, Prentice-Hall, 1975.

U. Rohde, "Low Noise Frequency Synthesizers Using Fractional N Phase-Locked Loops," *MSN*, January/February 1981.

E.R. Schineller, et al, "Development of a Fast Hopping Frequency Synthesizer Utilizing a Fractional Division Phase Locked Loop," ITT Avionics Div., Nutley, New Jersey, Final Report to Avionics Lab, Wright-Patterson AFB (AFWAL-TR82-1093), September 1982.

D. Sunderland, et al, "COM/SOS Frequency Synthesizer LSI Circuit for Spread Spectrum Communications,"

IEEE J. of Solid State Ckts., Volume SC-19, No. 4, August 1984.

J. Tierney, et al, "A Digital Frequency Synthesizer," *IEEE Trans. Audio Electroacoust.*, Volume AU-19, p. 48, March 1971.

C.E. Wheatley, III and D.E. Phillips, "Spurious Suppression in Direct Digital Synthesizers," *Proc. 35th Annual Frequency Control Symposium*, Philadelphia, May 1981.

C.E. Wheatley, III, "Digital Frequency Synthesizer with Random Jittering for Reducing Discrete Spectral Spurs," United States Patent 4,410,954, 18 October 1983.

Miscellaneous

M. Fashano and A. Strodbeck, "Communications System Simulation and Analysis with SYSTID," *IEEE on Selected Areas in Communications*, p. 8, Volume SAC-2, No. 1, June 1984.

A.V. Oppenheim, *Digital Signal Processing*, Prentice-Hall, 1975.

S.M. Selby, Ed, *CRC Standard Mathematical Tables*, 22nd Edition CRC Press, 1974.

ARMY FREQUENCY AGILE SYNTHESIZER PROGRAM

A. Bramble and J. Kesperis

US Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, New Jersey 07703-5000

Abstract

Objective of ET&DL Synthesizer Program: The Army has had several synthesizer developments over the last 28 years. However, the present program began in 1981 and differs from past efforts in that the major emphasis is directed toward the design of frequency agile synthesizers with multiple applications. The program objective is to retire the risk and lower the cost, size of synthesizer modules developed for secure Army communication equipment which operates in the HF through UHF bands.

Present ET&DL synthesizer Programs: Since March 1981, RCA has been under contract to ET&DL to develop a multirange, multiuse, agile frequency synthesizer CMOS/SOS chip set. The program defined present and future synthesizer requirements of Army communications equipment, studied capabilities of known synthesizer techniques, developed circuit designs to meet the requirements and fabricated devices and modules to prove the designs. This effort is nearing completion and synthesizer modules incorporating the chip set are being used in prototype systems.

A recent effort initiated as a logical follow-on to the RCA program is now underway as a joint program with DARPA and Rockwell Corporation. This study/development will utilize GaAs technology to extend the capabilities of Direct Digital techniques and realize monolithic synthesizer devices capable of wide bandwidth, fast switching and close channel spacing.

ARMY FREQUENCY AGILE SYNTHESIZER PROGRAM

Background - The present Electronic Technology and Devices Laboratory (ETDL) program was initiated in 1981 with the objective of lowering the risk and cost of synthesizer modules developed for Army equipment operating in the HF through low UHF bands. (Figure 1)

BACKGROUND

**PRESENT ET&DL SYNTHESIZER PROGRAM
INITIATED IN 1981**

- GROWING REQUIREMENT FOR SECURE COMMUNICATIONS
 - HIGH DEVELOPMENT AND LOGISTICS COST OF CUSTOM DESIGNS FOR EACH NEW SYSTEM
 - SIMILARITY OF REQUIREMENTS
 - AVAILABILITY OF CMOS/SOS AND GaAs TECHNOLOGIES
 - FEASIBILITY OF ALTERNATE CIRCUIT DESIGNS: DIRECT DIGITAL
- IS A MULTI-USE LSI SYNTHESIZER DESIGN FEASIBLE?

Increasing requirements for secure communications, scanning receivers and survivable surveillance equipment had resulted in expanded interest in frequency hopping capability. It was also apparent that a very significant percentage of system development cost resulted from the design and development of one-of-a-kind synthesizers for each new

system despite similarities in electrical operating characteristics. The increasing availability of the high speed capabilities of CMOS/SOS and GaAs had opened the door to new architectures and combinations of circuit techniques which had previously not been available and had restricted direct digital techniques to very low frequencies.

Army System Needs - Army systems requiring frequency agility fall generally into three categories: Communications, Electronic and Signal Warfare and Radar. Figure 2 outlines the direction that Communications systems requirements are headed.

SYSTEM REQUIREMENTS

COMMUNICATIONS

- MULTI FUNCTION SYSTEMS
- MULTI BAND PORTABLE RADIOS
- SPREAD SPECTRUM — FREQUENCY AGILE
- SHIFT TOWARD ALL DIGITAL IMPLEMENTATION OF ANALOG FUNCTIONS
- PROGRAMMABLE, ADAPTABLE TO JAMMING ENVIRONMENT

Near term requirements for man-portable HF and UHF radios are being addressed by PRC-104 upgrades and the SINCGARS development. Expected future capabilities of Electronic Warfare and Radar systems are also outlined in Figures 3 and 4.

SYSTEM REQUIREMENTS

ELECTRONIC WARFARE

- MULTI FUNCTION — COMBINED COMINT AND SIGINT
- RAPID SWITCHING
- NARROW FREQUENCY INCREMENTS
- LOW NOISE FLOOR AND SPURIOUS OUTPUT TO INCREASE PROBABILITY OF DETECTION
- LIGHT WEIGHT, LOW POWER FOR AIRBORNE APPLICATIONS
- LOW COST EXPENDABLE, ADAPTABLE JAMMERS

SYSTEM REQUIREMENTS

RADAR

- MULTI FUNCTION SYSTEMS-DETECTION, CLASSIFICATION, IDENTIFICATION
- FREQUENCY AGILITY
 - ANTI-JAM
 - LOW PROBABILITY OF DETECTION
 - PROVIDE INFORMATION OF TARGET PHYSICAL LENGTH
- RAPID FREQUENCY SWITCHING
- LOW CLOSE-IN PHASE NOISE
- OCTAVE BANDWIDTH SYNTHESIZERS FOR WIDE SPREAD SPECTRUM
- LOWER SIZE, WEIGHT, POWER FOR AIRBORNE APPLICATIONS

The obvious common requirement for all future systems is multi-function capability and frequency agility. Both of these requirements point toward increased use of digital circuitry. The required multiplication of system complexity dictates that cost and size will become important factors in determining producibility.

Army Synthesizer Program - There is an on-going effort in ETDL to provide devices and subsystems to upgrade present equipment capabilities and to provide new technologies and circuits for future needs. Figure 5 outlines the present program to provide for near and far term development of frequency agile synthesizers.

ARMY SYNTHESIZER DEVELOPMENT PROGRAM

PROGRAM (62 R&D)	CONTRACTOR	PERIOD	FUNDING	AGENCY
CMOS/SOS SYNTHESIZERS	RCA	81-86	ARMY	ARMY
GaAs ECCM CIRCUITS	ROCKWELL	82-85	ARMY/ DARPA	AF
GaAs DDS ARCHITECTURES	HUGHES	85-86	ARMY	ARMY
GaAs DDS ARCHITECTURES	HARRIS	85-86	ARMY	ARMY
HEMT TECH FOR ECCM	AT&T BELL LABS	84-87	ARMY	ARMY
DDS SYSTEM DEVELOPMENT	PLANNED	87-89	AI	ARMY
HEMT DDS SUBSYSTEMS	PLANNED	87-89	ARMY	ARMY

The CMOS/SOS program initiated in 1981 has been very successful although not all of the program goals were met. A part of the effort which was very successful was the development of a DDS Arithmetic Synthesizer. This development coupled with the results of the Rockwell MESFET contract which demonstrated the feasibility of fabricating GaAs MSI devices by producing 1 GHz, 1 μ m, 4 and 6 bit shift registers using the technology developed for the DARPA Pilot Line program indicated that further effort on DDS development was warranted. Based on successes of the RCA and Rockwell contracts two parallel studies were begun in 1985 with Hughes and Harris to investigate conventional and unconventional DDS architectures. A following effort is planned for 1987 to implement a selected DDS design based on the study results in GaAs MESFET technology. ETDL also is investigating HEMT technology for use in implementing DDS devices. A contract began in 1984 with AT&T Bell Labs demonstrating 7-10 GHz prescalers and 2-4 GHz, 16 bit

accumulators. A following contract is planned for 1987 to fabricate high speed HEMT DAC's and ROM's.

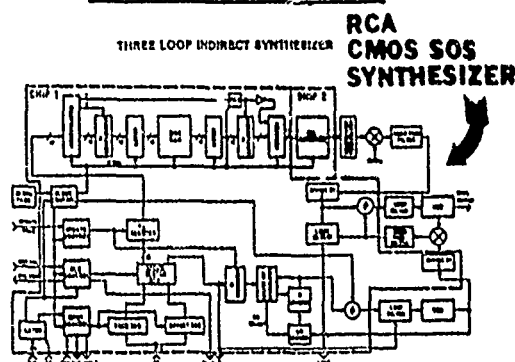
The near term requirements are being addressed by a program which began in 1981 with RCA for the development of a multi-usage frequency agile, wideband, CMOS/SOS synthesizer chip set. The general goals of the program are shown in Figure 6.

TECHNOLOGY FOR ANTI-JAM COMMUNICATIONS CMOS/SOS FREQUENCY SYNTHESIZER

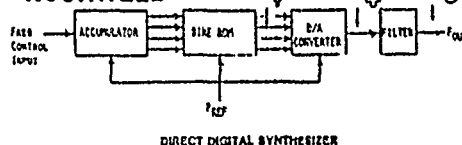
- WIDE APPLICATION: HF, VHF, UHF Communications and Data Link applications.
- FAST FREQUENCY HOPPING: Up to 5KHz hopping rate.
- WIDE BANDWIDTH: 30 to 470 MHz output frequency.
- SMALL STEP SIZE: Down to 25Hz channel spacing.
- LOW COST VLSI technology. Reduced applications engineering, reduced system logistics cost.
- USABLE: Users handbooks, application notes, computer programs available to reduce applications effort.

It was necessary to expand the chip set from two to three devices by developing a phase detector/reference divider to reduce the synthesizer output spurs resulting from crosstalk on the arithmetic chip. Maximum operating frequency exceeds 500 MHz, however the hopping rate achieved was lower than the objective shown. A block diagram of the synthesizer is shown at the top of Figure 7.

UHF SIGNAL PROCESSING



CIRCUITS FOR 200 MHZ GALLIUM ARSENIDE SYNTHESIZER BY ROCKWELL



The design is similar to the standard coarse loop, fine loop and summing loop architecture but in this design the fine loop has been replaced by a direct digital arithmetic synthesizer operating over the 1.25 to 2.25 MHz range. The control section of the arithmetic chip contains all of the elements required to accept either eight decade serial or parallel eight

bit byte frequency word input commands T/R and 1K offset control. A direct frequency control input of the direct digital synthesizer is also provided. This program is nearing completion and has been very successful. Synthesizer modules incorporating the CMOS/SOS chip set have been delivered to Hazeltine for use in prototype HF SNAP models. Samples of the devices have been provided to ITT and Magnavox for possible use in on-going military equipment developments. The RCA synthesizer is compared in Figure 7 to a conventional DDS implemented in GaAs digital LSI. A DDS operating at a clock rate of 2 GHZ with an output in the 100 to 200 MHZ range with small output increments and high switching speeds would meet 75% of the Army's synthesizer requirements if the spurious output requirements could be met.

The Army is involved in three generations of synthesizer technology development as shown in Figures 8, 9 and 10, which seek to develop anti-jam synthesizers with output frequencies up to 500 MHZ.

ANTI-JAM SYNTHESIZER TECHNOLOGY DEVELOPMENT

THREE GENERATIONS OF TECHNOLOGY ARE IN DEVELOPMENT:

1. SILICON CMOS/SOS.....

- INDIRECT DIGITAL SYNTHESIZERS INCORPORATING PHASE LOCKED LOOPS AND ARITHMETIC (DIGITAL) SYNTHESIZERS
- FREQUENCY AGILE OPERATION TO LOWER UHF
- CHIP-SETS AND MODULES HAVE BEEN DELIVERED TO SYSTEM DEVELOPERS FOR ENGINEERING EVALUATION

2 GALLIUM ARSENIDE MESFET TECHNOLOGY.....

- DIRECT DIGITAL, FAST SWITCHING FREQUENCY AGILE SYNTHESIZERS
- ARCHITECTURES FOR OPERATION TO 100, 225 and 500 MHz ARE IN DESIGN
- MSI DDS CIRCUIT PROTOTYPES HAVE BEEN TESTED AT 1 GHz
- A FULLY MATURE LSI/VLSI TECHNOLOGY IS NOT YET AVAILABLE

3. HIGH ELECTRON MOBILITY TRANSISTOR TECHNOLOGY.....

- MULTI-GIGAHERTZ LSI TECHNOLOGY IN EARLY STAGES OF DEVELOPMENT
- LSI 8-16 BIT ACCUMULATORS FOR DDS SYSTEMS ARE IN DESIGN
- OFFERS THE BEST SPEED-POWER AT HIGH CLOCK-RATES, AND OPERATION AT HIGH CLOCK TO OUTPUT FREQUENCY RATIO FOR ENHANCED DDS SPECTRAL PURITY

This involves two generations of advanced device technology beyond CMOS/SOS. The first, as shown in Figure 9, is the MESFET GaAs technology which offers speed advantages over CMOS/SOS and the second, Figure 10, is the still faster High Electron Mobility Transistor (HEMT) technology which is rapidly emerging but is longer range in application.

The present Hughes and Harris programs are studying a design matrix, outlined in Figure 11.

CONVENTIONAL DDS SYNTHESIZER
COMPONENT REQUIREMENTS VS PERFORMANCE

MAX OUTPUT (MHz)	MIN CLOCK (MHz)	FREQ RES (Hz)	ACCUM WIDTH (bits)	-55 dBc		-80 dBc	
				D/A (bits)	ROM WIDTH	D/A (bits)	ROM WIDTH
100	400	10	26	10	10	14	14
		1K	19	10	10	14	14
		1M	9	10	10	14	14
225	1000	10	27	10	10	14	14
		1K	20	10	10	14	14
		1M	10	10	10	14	14
500	2000	10	28	10	10	14	14
		1K	21	10	10	14	14
		1M	11	10	10	14	14

Both conventional and innovative architectures are being examined and simulated with respect to major tradeoffs between performance and technological realizability. A major motivation for innovative architectures is to remove the burden of high resolution D/A converters and ROMs operating at DDS system clock speeds as they must in the conventional DDS implemented with an accumulator, ROM and D/A. At this point it appears that HF to lower VHF DDS GaAs MESFET conventional synthesizers can be realized no earlier than three years, and only with further development.

Figure 12 outlines supporting technology programs in which the Army is deeply involved.

**ARMY SYNTHESIZER DEVELOPMENT PROGRAM
(SUPPORTING PROGRAMS)**

<u>PROGRAM</u>	<u>CONTRACTOR</u>	<u>PERIOD</u>	<u>FUNDING</u>	<u>AGENCY</u>
Si & GaAs SUBMICROMETER PROCESS SIMULATORS	STANFORD	86-88	DARPA/ VHSIC	ARMY
GaAs PILOT LINE DEVELOPMENT	ROCKWELL HONEYWELL McDONNELL	84-87	DARPA Tri-Service Monitored	AF
R&D ON GaAs DESIGN, CAD, & SIMULATION TECHNOLOGY	MAYO CLINIC	83-87	DARPA ARMY COTR	DSSA

A program beginning in 1986 with Stanford University, jointly funded by DARPA and the DOD VHSIC Office, will develop CAD process simulators for Silicon and GaAs MESFET technology. A Tri-Service effort with Rockwell, Honeywell and McDonnell Douglas, will develop a MESFET pilot line to solve producibility problems, increase yield enhancement and increase process line throughput. The final effort is an on-going program with the Mayo Clinic for R&D on GaAs MESFET design, gate array personalization and processor CAD and simulation technology. The above technology supporting programs comprise a \$5M effort over a five year period.

SOS FREQUENCY SYNTHESIZER DEVELOPMENT

D. P. O'Rourke

RCA Communication and Information Systems Division
Camden, New Jersey 08102

RCA Advanced Technology Laboratories and Communication and Information Systems Division, under contract to U.S. Army ERADCOM, have defined a synthesizer architecture and have developed a three LSI chip set in CMOS/SOS set aimed at meeting the military's present and future systems requirements. The major goal of the program was the development of new synthesizer circuit designs and/or refinement of known techniques, such as direct, direct digital and indirect frequency synthesis to formulate a multi-range synthesizer that can be versatile for various applications, adaptable to changes in system specifications over time and in which a majority of the circuit functions can be implemented in monolithic integrated circuit form. In a continuation of the original contract, computer programs were developed to provide a Computer-Aided Design package consisting of translation loop PLL design and analysis and CAD for voltage controlled oscillators.

Requirements by the military for secure communications using spread spectrum techniques have intensified the need for low-cost, high-purity, fast switching frequency synthesizers, especially for the frequency-hop form of spread spectrum. Additional requirements include narrow channel spacing for efficient spectrum utilization, low noise to prevent desensitization of co-located equipments, and microprocessor controllability.

RCA Advanced Technology Laboratories and Communication and Information Systems Division, under contract to U.S. Army ERADCOM, have defined a synthesizer architecture and have developed a three LSI chip set aimed at meeting these synthesizer requirements. The chip set integrates the digital and some analog portions of a two-loop synthesizer design. The use of CMOS/SOS VLSI offers the additional advantages of low power consumption, high reliability, small size, and reduced weight, while maintaining high performance levels.

The synthesizer architecture combines direct digital techniques for narrow channel spacing with phase-locked loop techniques in a two loop synthesizer. The architecture and its associated chip set are used in HF, VHF and UHF synthesizer designs.

The LSI chip set includes a synthesizer array which contains the bulk of the digital circuitry - phase detectors, digital synthesizer, controller, dividers and counters, a 12-bit D/A converter and a separate phase detector array.

In a continuation of the original contract, computer programs are being developed to provide a computer-aided design package consisting of

translation loop PLL design and analysis and CAD for voltage controlled oscillators.

Synthesizer Development

The major goal of the program was the development of new synthesizer circuit designs and/or refinement of known techniques, such as direct, direct digital and indirect frequency synthesis to formulate a multi-range synthesizer in which a majority of the circuit functions can be implemented in monolithic integrated circuit form. In addition, the synthesizer configuration must be able to fulfill the military's present and future systems requirements, be versatile for various applications and be adaptable to changes in system specifications over time. By using advanced technology CMOS/SOS fabrication, the synthesizer will be small in size, low in power consumption and radiation tolerant, making it an attractive candidate for technology insertion to upgrade present systems as well as providing an available low-cost, high-performance synthesizer for future designs.

The RCA design, which covers the 30 to 500 MHz range, uses a translation loop approach whereby the output of a low-frequency, fine-frequency-stepping generator is translated up to the desired output frequency. To provide small frequency steps, RCA employs a Direct Digital Synthesizer. By using the DDS, which can be implemented in LSI, numerous bulky analog components can be eliminated without significantly sacrificing spectral purity and settling time. A general block diagram of the synthesizer configuration is shown in figure 1. Functions contained on one or more LSI circuits are enclosed within dashed line. Several

synthesizer configurations covering wide frequency bands from 30 to 500 MHz were studied using the same VLSI chips and changing only the VCOs, a minimum number of external components, and selecting the proper reference and feedback dividers in the PLL configuration.

Synthesizer Array (TA12779)

The bulk of the synthesizer digital circuitry is contained on the Synthesizer Array. The circuitry consists of input interface and control registers, an ALU for frequency decoding, two phase detectors for both PLLs, variable and fixed frequency dividers for the reference and VCO and an out-of-lock detector. The device is fabricated in CMOS/SOS and packaged in a 64-pin ceramic leadless chip carrier (LCC) or 64-pin, Pin Grid Array (PGA).

Direct Digital Synthesizer - The frequency generating portions of the synthesizer array start with the DDS. The DDS consists of a digital accumulator which adds the contents of an input frequency register to the contents of the accumulator at every transition of the input clock signal. The accumulator number grows linearly until the capacity of the accumulator is exceeded. At this point, the accumulator overflows and begins to linearly increase again, in a sawtooth manner, as further register outputs are accumulated. The total capacity of the accumulator represents 360 degrees of phase change or one full cycle. The rate of overflow is a measure of frequency and can be controlled by changing the number in the frequency register called the Frequency Control Word (F_N). This control word essentially represents a step in phase, and the parallel output of the accumulator is, therefore, the digitally encoded phase of the desired output frequency. To generate a sinewave output, quadrant logic and a sine function generator in a Read-Only Memory (ROM) are placed between the accumulator and D/A Converter.

The ROM typically stores only one-quarter of the sinewave. The quadrant logic converts the sawtooth variation of the accumulator output into a triangular varying output prior to entering the ROM. At the output of the ROM, similar logic either passes or complements the output data to the D/A Converter to generate the second 180 degrees and, thus, the bipolar nature of the sinewave.

The numerical output of the sine function generator or quadrant logic is fed to the Digital-to-Analog Converter (DAC) which produces a sinusoidal output waveform. This waveform consists of sine weighted staircase sampling levels which can be removed with an output lowpass filter producing a clean sinewave output.

The output frequency of the synthesizer (F_O)

is related to the number in the register (Frequency Control Word, F_N), the accumulator capacity (N_C) and the clock frequency (F_C) through the expression:

$$F_O = ((F_N/N_C) \times F_C)$$

As indicated from this formula, there are three variables involved in calculating the output frequency. The maximum frequency is restricted to a fraction of the input clock frequency, and the maximum input clock frequency is limited by the speed of the logic used in the accumulator, the ROM and primarily the D/A Converter.

Complete Synthesizer - The overall synthesizer architecture consists of a high frequency coarse stepping divide-by-N/prescaler PLL, a low frequency Direct Digital Synthesizer (DDS) to interpolate between the wide-steps of the divide-by-N PLL and a translation PLL to sum together these two effects.

12-BIT D/A Converter

The D/A is a key component in the current DDS design. Principally, it determines the maximum allowable clock rate and, therefore, the maximum output frequency and the overall spurious level. Initial investigations centered on an analysis of commercially available D/A converters. It was determined that the best available devices would not meet the requirements for a low spurious, low power, high frequency DDS. However, from these tests and computer analysis, guidelines for the development of a custom D/A converter were conceived. The D/A converter was fabricated as a separate LSI and not integrated with the other functions of the DDS due to the special performance characteristics required and the increased possibility of cross-coupling analog and digital signals. The D/A converter uses a novel electrical trimming technique which permits high clock rates but with low glitch energy on the output waveform. The D/A performance exceeds the DDS requirement for a low frequency generator (2-3 MHz) with very fine frequency resolution, low glitch energy and 12-bit resolution. The device is fabricated in CMOS/SOS and packaged in a 28-pin ceramic dual-inline.

Phase Detector LSI

The phase detector array is constructed to interface directly in parallel with the control inputs of the Synthesizer Array. The circuit consists of a dual flip-flop phase frequency detector identical to the phase detectors on the Synthesizer Array and a variable reference divider. This external detector is required for synthesizer applications where low noise, low spurious and high isolation is required. It is fabricated in CMOS/SOS and packaged in a 20-pin ceramic dual-inline.

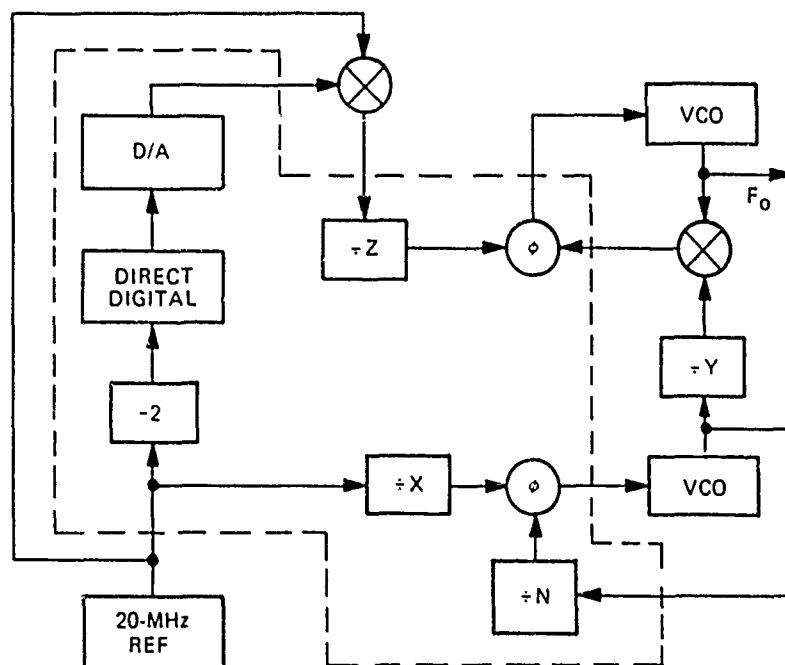
CAD Software

As part of a continuation of the original SOS Frequency Synthesizer contract, RCA is developing a computer-aided design program to complement the LSI hardware. The program will be specific to the LSI chip set and translation loop synthesizer architecture. The individual PLL design will center on Type II, third order loops using current pumps and passive filters. It has generally been found this configuration provides the most flexibility for a wide variation of performance parameters and applications for a fixed baseline architecture. An additional program calculates initial component values for a Colpitts VCO configuration over a user specified frequency range.

Conclusion

Military and Government communications equipment require increasingly complex multi-channel, frequency-hopping, low noise modes of operation in order to meet objectives for secure voice and

data communication. The frequency synthesizer is a major component in the determination of the ultimate performance of these new equipment applications. A list of the important performance parameters includes narrow channel spacing, low noise, frequency agility, low power, low cost, spectral purity, small size, and micro-processor controllability. RCA has developed a three LSI synthesizer chip set which integrates a majority of the digital logic of a translation loop synthesizer. The unique translation loop design combines the effects of a fine frequency stepping Direct Digital Synthesizer with a coarse incrementing PLL. The result is a baseline configuration adaptable to various synthesizer applications for future synthesizer requirements or upgrading of present communications equipment. The call for low cost and high performance circuits has been answered with LSI circuits, and the application of this technology to frequency synthesis will have an important impact in this field. The use of CMOS/SOS will enhance synthesizer designs with high speed, low power, low costs and micro-processor control capability.



FUNCTIONS CONTAINED ON ONE OR MORE LSIs
ENCLOSED WITHIN DASHED LINE

FIGURE 1. BLOCK DIAGRAM —
TRANSLATION LOOP SYNTHESIZER

Performance Analysis of the Numerically Controlled Oscillator

Elizabeth C. Kisenwether, Member, IEEE

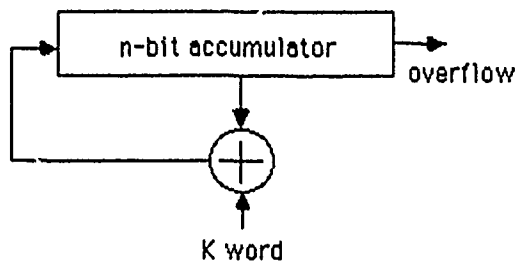
William C. Troxell

Summary

With the advances in high-speed VLSI and VHSIC logic, use of digital logic in high speed oscillator design is now possible. One design approach is the numerically controlled oscillator (NCO), where continuous summations of step size K are made to the previous accumulator value. Overflow of the accumulator results in an output at the desired periodic rate. This paper addresses operation of the direct synthesis NCO including frequency determination and control, phase adjustment, and phase jitter performance.

Numerically Controlled Oscillator Definition

Figure 1 shows a simplified block diagram of a direct synthesis NCO. This method of direct digital synthesis is often referred to as phase accumulation in the literature, but this terminology is avoided in this paper because intentional phase adjustments, called $\Delta\theta$, are later discussed in detail.



Numerically Controlled Oscillator

Figure 1

The accumulator is incremented in step size K to a maximum accumulator count of 2^R , where R is the number of accumulator bits. In general, when the accumulator overflows a residual count remains in the accumulator, and is the beginning count in the accumulator for the next summation cycle. K -step size summations occur continuously at the reference clock rate of f_{ref} . The basic NCO operation is shown in Figure 2.

The actual NCO output frequency can be obtained a number of ways. One method is to use the most significant bits (msb) out of the accumulator as address information for a sine or cosine lookup table in

RDM (read-only memory). The RDM table performs the digital translation from NCO count value to sinewave magnitudes, which are then sent to a follow on digital-to-analog (D/A) converter and low pass filter. The number of accumulator bits, n , determines the sinewave quantization. the ± 1 amplitude range of the output sinewave is defined in increments of 2^n .

A second method of NCO signal generation is to use the msb of the accumulator, providing a squarewave frequency defined as:

$$f_{NCO} = 0, \quad 0 \leq \text{msb} < (1/2) 2^R$$

$$f_{NCO} = 1, \quad (1/2) 2^R \leq \text{msb} \leq 2^R$$

R = accumulator register length

The accumulator msb changes polarity at an approximately 50% duty cycle, and can be used as the system clock or as a reference clock for a phase locked loop (PLL) section. The PLL generates the final output frequency, f_{PLL} . A PLL section is typically used to upconvert the NCO output to a higher frequency than possible by direct digital synthesis.

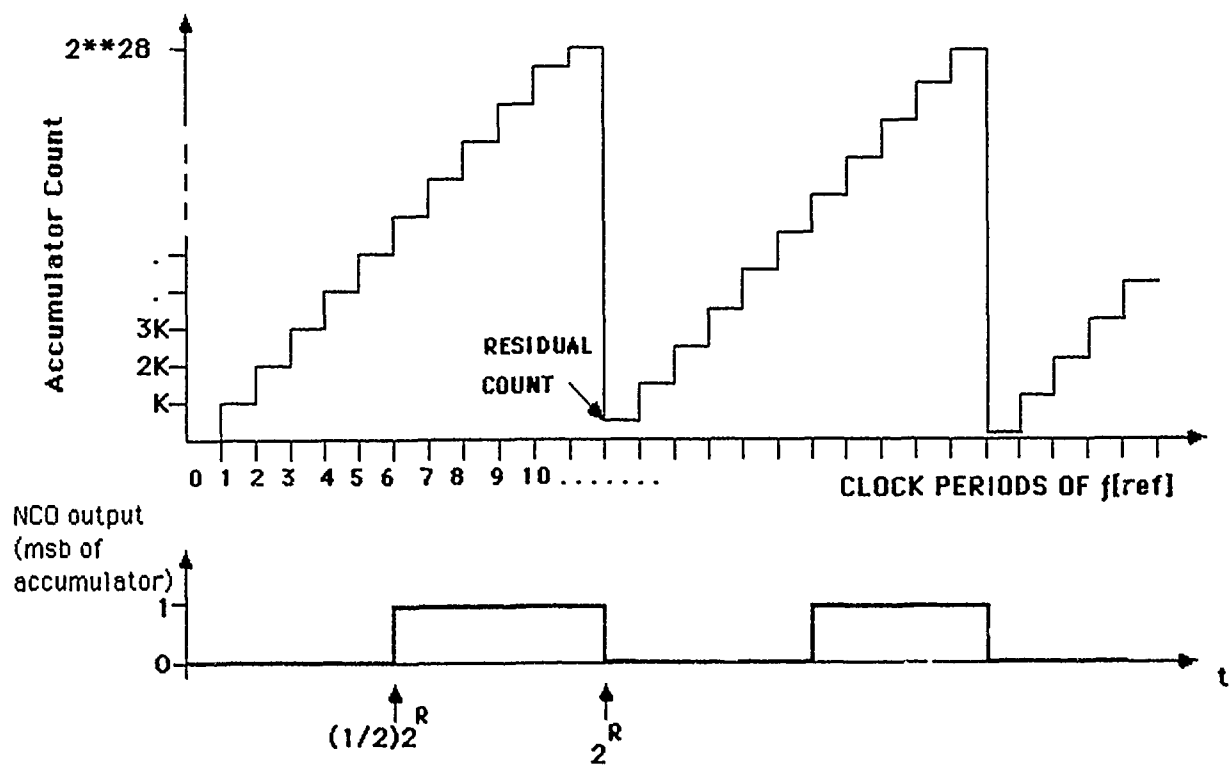
This second NCO approach of an Accumulator/PLL pair for oscillator signal generation is the topic of analysis in this paper.

The NCO frequency, measured at the msb of the digital accumulator, is directly related to the accumulator step size K , accumulator register length R , and the reference frequency f_{ref} :

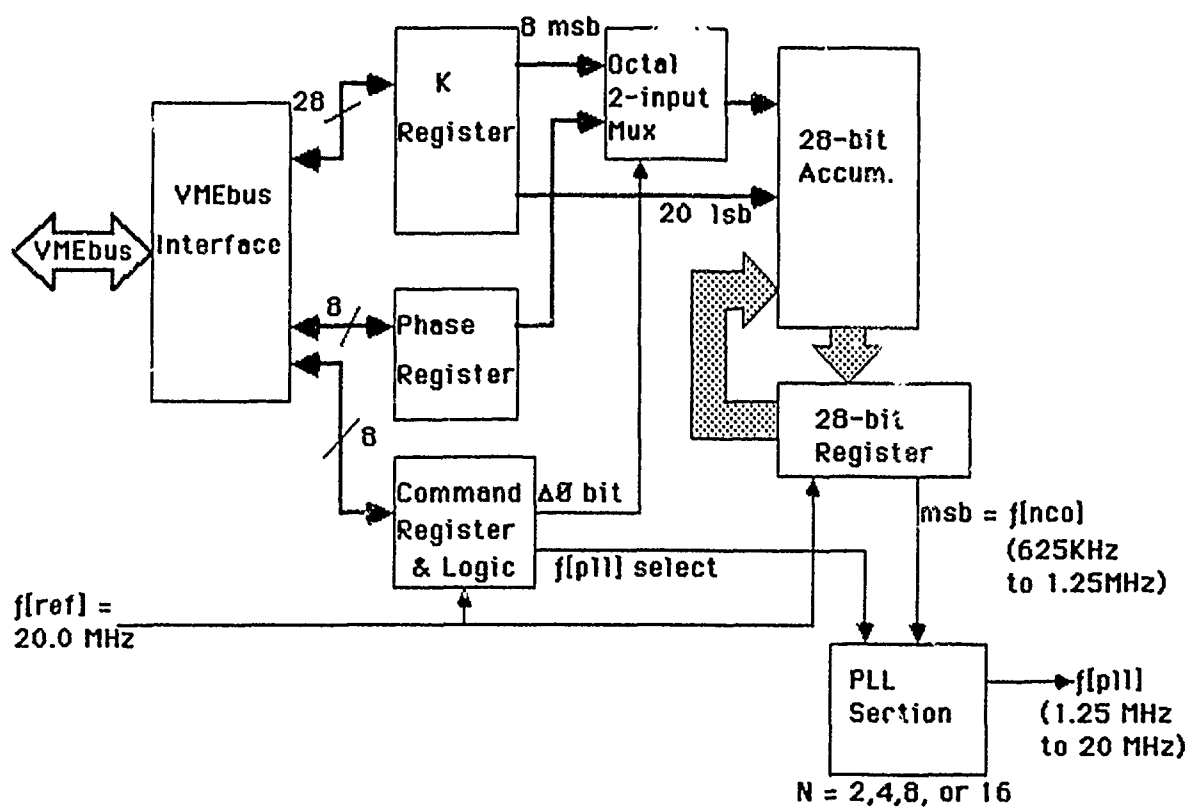
$$f_{NCO} = \frac{K \times f_{ref}}{2^R} \quad (1)$$

Equation (1) is the fundamental equation for NCO operation. Typical values for actual implementation in hardware are $R = 28$, and $f_{ref} = 20.0 \text{ MHz}$. These values give the NCO frequency a resolution of 0.0745 Hz (obtained by setting $K=1$ in Equation (1)). One implementation of the NCO is shown in the block diagram of Figure 3.

This paper analyzes four areas of NCO performance: NCO frequency bounds, K step roundoff effects, phase change commands, and NCO phase jitter.



Graphical representation of NCO operation
Figure 2



NCO Board Block Diagram
Figure 3

Upper Bounds on f_{NCO}

The NCO is designed to provide high resolution frequency selection to the subhertz level by division of the reference frequency f_{ref} . The NCO frequency, f_{NCO} , is derived from the msb of the accumulator output. For a large K step size (relative to the maximum accumulator count 2^R), the accumulator will overflow rapidly, yielding a f_{NCO} which is fast (i.e., a large fraction of the reference frequency f_{ref}). For a small step size K, thousands of repeated accumulations of size K must occur for accumulator overflow, giving a slow NCO output frequency.

It is the large step size of K (and resulting high f_{NCO}) which bounds NCO frequency output quality. In general, a large K-step value results in large phase jitter in f_{NCO} because the reference frequency period is only slightly shorter than the NCO frequency period. The effects of K step size are addressed in more detail in a following section entitled Phase Jitter Effects.

Small values of K result in f_{NCO} much lower than the reference frequency f_{ref} , so the period of the NCO signal, T_{NCO} , is many orders of magnitude longer than the reference signal period T_{ref} . For example, with $K=13,421$, the resulting NCO frequency is 1KHz (with $R=28$ and $f_{ref} = 20.0$ MHz). For this case, $T_{NCO} = 20,000 T_{ref}$.

Errors due to K roundoff

From Equation (1), the K-step value to generate a particular f_{NCO} is defined as:

$$K = \frac{f_{NCO} \times 2^R}{f_{ref}} \quad (2)$$

In general, the calculated K will be non-integer for the desired f_{NCO} . Any rounding of the K-step value to the closest integer value results in error between the desired f_{NCO} and the actual f_{NCO} .

The maximum roundoff in K is ± 0.5 to reach to closest integer value. The maximum error in f_{NCO} due to the roundoff ΔK is Δf_{NCO} .

$$\Delta f_{NCO} = \frac{\Delta K \times f_{ref}}{2^R} \quad (3)$$

Using $f_{ref} = 20$ MHz and $R=28$, $\Delta f_{NCO} = 0.0373$ Hz

As mentioned earlier, the f_{NCO} can be used as the reference frequency for a phase locked loop, which provides upconversion to the required output frequency range. The roundoff error from the accumulator section is multiplied by the upconversion factor N. The amount of upconversion used is bounded by the amount of error that can be tolerated in the final upconverted frequency f_{PLL} . The larger the upconversion factor N, the larger the error in f_{PLL} due to roundoff in K. Additionally, divider size has a direct effect on PLL phase jitter. For applications where frequency accuracy is important, the upconversion factor must be kept to a minimum. For the NCO design shown in block diagram Figure 2, the maximum value of $N=16$ was used, yielding worst case error between desired f_{PLL} and actual f_{PLL} of $\Delta f_{PLL} = 16 \times 0.0373$ Hz = 0.5960 Hz using the conditions of Equation (3).

If high final output frequencies are required, such as $f_{PLL} > 40$ MHz, an upconversion factor $N > 16$ can be used, with according increases in Δf_{PLL} .

Phase commands and $\Delta\theta$ performance

Referring to Figure 2, the method used for changing phase of the f_{NCO} (and the phase of f_{PLL}) is presented. For normal free-run operation, where no phase changes are commanded, the full 28 bits of the K-step are used in the accumulation process. The octal 8-bit mux of Figure 2 is set to send the 8 msb of the K-step value to the accumulator.

When a phase change in f_{NCO} or f_{PLL} is desired, the following steps occur:

1. The operator enters the desired phase change $\pm\Delta\theta$ (in degrees) on the VMEbus system keyboard.
2. Software determines the amount that K must be changed to achieve the $\Delta\theta^\circ$ phase adjust.

$$\Delta K\theta = \frac{\Delta\theta \times 2^R}{360^\circ} \quad (4)$$

3. A negative phase shift ($-\Delta\theta$) is accomplished by decreasing the value of K by $\Delta K\theta$ for one period of the reference clock f_{ref} . Since the K-step size is reduced in magnitude, more time is required to reach maximum accumulator count of 2^R . A positive phase shift ($+\Delta\theta$) is accomplished by increasing the K-step size by $\Delta K\theta$ for one period of the f_{ref} clock period.

4. Software calculates the $K \pm \Delta K\theta$ value, converts it to

binary, and sends the 8 msb of the $K \pm \Delta K$ over the VMEbus to the Phase register on the NCO board.

5 The $\Delta\theta$ command bit in the Command register is toggled from a 0 to a 1 under software command to execute the phase adjust. With the change in the $\Delta\theta$ command bit, the 8-bit mux selects the eight bits of Phase register instead of the 8 msb of the K-step. For one cycle of accumulation (i.e., on period of f_{ref}). The altered step size $K \pm \Delta K$ is used in the accumulation process, then the command logic reselects the 8 msb of the original K-step value. The NCO signal returns to the original T_{NCO} period automatically, so the NCO output frequency is unchanged, except for a change in phase.

Since only the eight bits of Phase register are available in this approach, lack of control over the least significant (R-8) bits of the K-step value results in an error between the desired $\Delta\theta^\circ$ and the actual $\Delta\theta^\circ$. Phase error θ_e is calculated to be.

$$\theta_e = \frac{2^{(R-8)} \times 360^\circ}{2^R} \quad (5)$$

For $R=28$, the low 20 bits of K are not changed in a $\Delta\theta$ operation, and $\theta_e = 1.4^\circ$.

It should be noted that the magnitude of the K-step value is a factor in NCO $\Delta\theta$ command selection. If $K < 2^{(R-8)}$, the 8 msb of the K are all zero. Since the phase change operation manipulates the only the 8 msb of K, it would appear that a slow f_{NCO} (small K-step values) would have poor $\Delta\theta$ command response. However, because K is small, many summations of K occur to reach maximum count of 2^R . For example, if $R=28$ and $K = 2^{19}$, this K value is summed 512 times per NCO period. Setting Phase register = 00000001 (bit $2^{20} = 1$), $K_{\Delta\theta} = 2K$, and maximum accumulator count of 2^{28} is reached one period of f_{ref} sooner. For this case $\Delta\theta^\circ = (1/512) \times 360^\circ = 0.7^\circ$.

Another question on NCO $\Delta\theta$ command accuracy involves the relative magnitude of T_{NCO} and T_{ref} . Define $M =$ number of summations of K to reach maximum count of 2^R . The question is whether $\Delta\theta^\circ$ is restricted to increments of $360^\circ/M$, or can the NCO signal be changed in phase by any amount. The answer is found by viewing the $\Delta\theta$ command as change of the residual count of the accumulator summation cycle

For $f_{ref} = 20.0$ MHz, $R = 28$, and $f_{NCO} = 1.25$ MHz, $M = 16$. Phase adjusts are accomplished by changing K to $k \pm \Delta K$

for one of the sixteen summations in the NCO period, as shown in Figure 3. Assume that without a phase adjustment, the residual count in the accumulator register would be X. With the $\Delta\theta$ change, the residual count is modified to $X \pm \Delta K$ (modulo 2^R). Therefore, the $\Delta\theta$ change can be viewed as an alteration in residual (overflow) magnitudes, and the smallest $\Delta\theta^\circ$ command is not restricted to integral values of $360^\circ/M$. The ΔK value used in determining $K_{\Delta\theta}$ defines the longterm phase shift obtained in f_{NCO} :

$$\Delta\theta^\circ = \frac{\Delta K \times 360^\circ}{2^R} \quad (6)$$

Phase Jitter Effects

Since the NCO is a repeated accumulation and overflow operation at the reference clock frequency f_{ref} , edge jitter of T_{ref} seconds will occur when the number of accumulations of K to reach maximum count changes on consecutive f_{NCO} cycles.

As defined in the previous section, M is the number of accumulations of step value K to reach maximum count 2^R .

$$M = \frac{2^R}{K} \quad (7)$$

In general, M is non-integer, which results in a varying number of summations of K required to obtain a count in the accumulator of 2^R or overflow. For example.

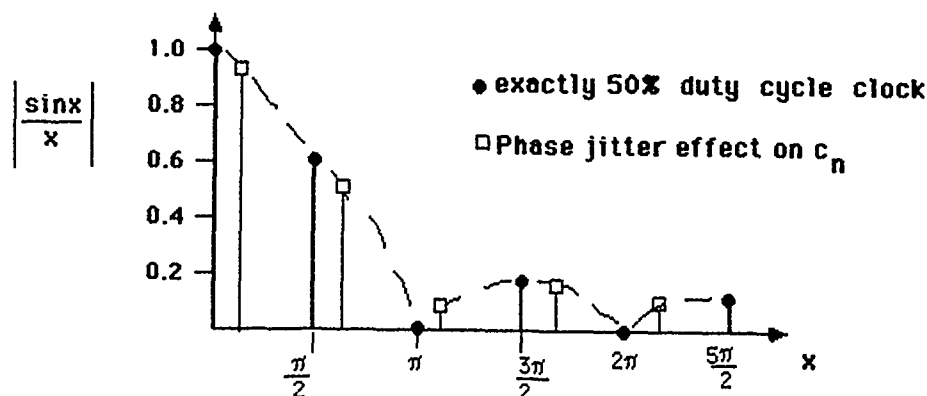
$$\begin{aligned} f_{NCO} &= 937,500 \text{ Hz} \\ K &= 12,582,913 \\ M &= 21.3333... \text{ for} \\ f_{ref} &= 20.0 \text{ MHz and } R = 28. \end{aligned}$$

The NCO accumulators operate by summing K for $M = 22$ times to reach overflow, then $M = 21$ times to reach overflow, then $M = 21$ times again. The resulting average value is $M = 21.3333...$, since the {22, 21, 21} pattern repeats endlessly until the accumulators are shut down, or a phase adjust is commanded.

The variations in the integer M values seen in the accumulator results in a 50 nsec jitter (one period of f_{ref}) for each change in M per accumulator cycle.

The phase jitter behavior in the NCO can be viewed a number of ways, expressing the phase jitter itself in the frequency domain, and developing a Discrete Fourier Transform description of M.

The effects of NCO phase jitter are shown in the frequency domain, as in Figure 4. For a perfect 50%



Effect of variations in M on Fourier Coefficient c_n

Figure 4

duty cycle pulse train, $t_0 = T/2$, and the Fourier coefficients are located at $n(\pi)/2$:

$$c_n = \frac{2At_0}{T} \frac{\sin(n\pi t_0/T)}{(n\pi t_0/T)} \quad (8)$$

For phase jitter of quantity T_{ref} , the Fourier coefficients are moved from $n\pi/2$ to $(n\pi/2 \pm n\pi(T_{ref}/T_{ncol}))$ at occurrence of the jitter.

$$c_n = 2A \left(\frac{1}{2} \pm [T_{ref}/T_{ncol}] \right) \times \frac{\sin \{n\pi[1/2 + [T_{ref}/T_{ncol}]]\}}{(n\pi[1/2 + [T_{ref}/T_{ncol}]])} \quad (9)$$

From Equation 8, it is evident that a short NCO period T_{ncol} results in larger spectral jitter. This derivation gives frequency domain justification for keeping K small. With the T_{ncol} long compared to the reference frequency period T_{ref} , jitter will still occur, but the span of the frequency jitter will be less.

A second phase jitter analysis gives additional insight into NCO performance. It has been shown that overflow of the phase accumulator does not always occur with the same number of clock cycles. Consequently, modification of the duty cycle or phase jitter occurs. This section quantifies the effect of the phase jitter.

The maximum accumulator value, or value required for overflow is related to clocking by

$$2^R = KM + X \quad (10)$$

where K = increment added to accumulator each clock cycle

M = integer number of clock cycles

X = accumulator residue after overflow

The value X is the cause of unequal clock cycle accumulation periods. If the accumulator size 2^R is evenly divisible by KM , there will be no residue X . Seldom, however, is this the case.

The number of clock pulses required to cause the accumulator to overflow is:

$$N_{80} = 2^R / \gcd(K, R) \quad (11)$$

\gcd = greatest common divisor

The number of output pulses, or in our case, output clock cycles for N_{80} input clock pulses is:

$$N_{op} = K / \gcd(K, X) \quad (12)$$

For these output cycles N_{op}

i occur after $n+1$ clock pulses

$[K/\gcd(K, d) - 1]$ occur after n clock pulses

$$\text{where: } i = X / \gcd(K, X) \quad (13)$$

The above formula predicts NCO output signal jitter, indicating that i cycles will have elongated (by one clock cycle) duty cycles while $[K/\gcd(K, d) - 1]$ will have exactly the correct duration M . The combination of normal and elongated cycles is periodic and dominates the phase jitter.

characteristic of the NCO. This period is given by N_{op} .

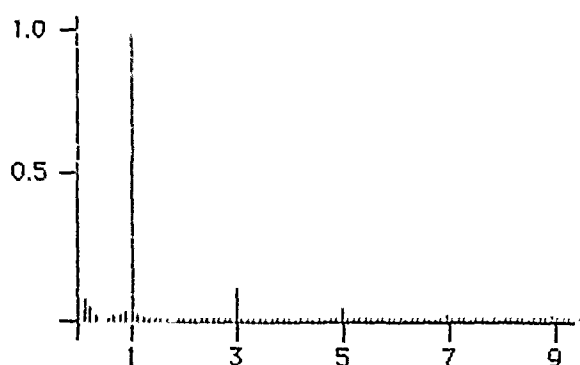
Note that this period describes the pattern of clock cycles required for accumulator overflow. As an example, consider the case where $R = 8$ and $K = 11$. Using the previously defined equations we find that:

$$M = 23 \quad X = 3 \\ N_{80} = 256 \quad N_{op} = 11 \quad i = 3$$

The number of input clock cycles required for each output cycle can be shown to be:

					Period							
	24	23	23	24	23	23	23	24	23	23	24	23
	24	23	23								

Figure 5 shows the spectrum of this periodic function. It was obtained by applying an unwrapped DFT to the previous periodic sequence.



Normalized and Linear NCO Output Spectrum
Figure 5

References

1. Tierney, J., Rader, C.M., and Gold, B., "A Digital Frequency Synthesizer". IEEE Trans. on Audio and Electronics; Vol. AU-19, No. 1, March 1971; 48 - 57.
2. Copper, H.W., "Why complicate frequency synthesis?". Electronic Design; Vol. 15, July 19, 1974; 80-84.

THE COHERENCE OF A RADAR MASTER OSCILLATOR.

R. D. Weglein, Hughes Aircraft Company,
Canoga Park, CA 91304

Abstract

The stability of a radar master oscillator is usually defined in terms of a) its ability to remain on frequency over a long time period, and b) the spectral line width of its carrier. The behavior of the oscillator in category b) is usually ascertained by the measurement of short term stability or phase (and amplitude) noise in the vicinity of the carrier. In pulse doppler applications, phase stability is required over a time period long compared to the radar signal round trip delay. When put in these terms, the delay interval during which the oscillator must remain coherent, is not readily available from direct measurements on the oscillator.

In this paper, a coherence interval is defined during which the rms phase between a perfect and a noisy oscillator changes by one radian. It is shown that the coherence interval may be estimated from the measured power spectrum. Examples are given to show that the suitability of a MM-wave source to a coherent radar may be estimated readily in this way. An experimental technique is proposed to determine the coherence interval.

Introduction

The short term stability of an oscillator is a well-defined parameter, by which the source's power spectrum in the neighborhood of its carrier is characterized. We tend to think of short term stability as a frequency domain phenomenon because that is the domain in which the power spectrum is most often acquired. This noise power spectrum is generally composed of two regions. In the immediate vicinity of the carrier frequency f_0 , phase noise is frequency-dependent and confined to the filter bandwidth of the frequency-determining resonator. Beyond this extends a relatively flat, frequency-independent "noise" floor. This somewhat idealized spectral plot, shown in Figure 1, will suffice for the purpose of this discussion. The intercept between the two regions is sometimes referred to as "corner" frequency, given by f_0/Q , where Q is the resonator's loaded Q . The experimental plot is obtained

by phase comparison of the "noisy" oscillator with a much "cleaner" (higher short-term stability) reference source at the same frequency. Alternately, two test oscillators, when available, may be phase compared with each other, with attendant increased system sensitivity.[1]

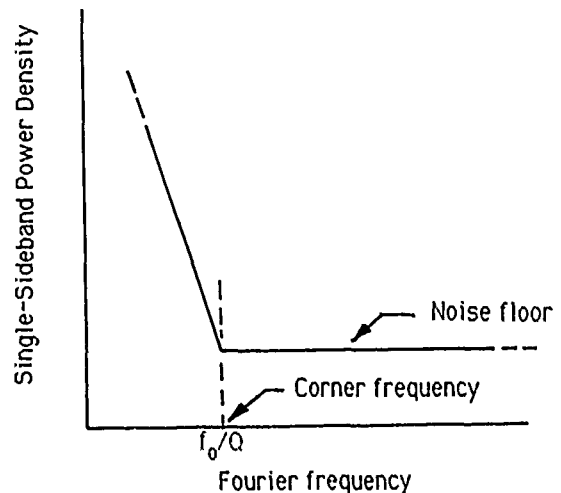


Figure 1. Simplified power spectral density of an oscillator showing the corner frequency f_0/Q .

A valid time-domain description exists also that makes use of the "Allen Variance" method, through which oscillator stability is sometimes characterized. [2] This latter view is more often called upon to define medium and long-term stability, such as frequency drift and aging in very stable references. Neither description is directly applicable to the coherence measurement of a radar master oscillator.

The spectral density description of Figure 1 is, however quite useful in a coherent (pulse-doppler, for example) missile radar. The doppler-shifted frequency in such a radar is a direct measure of the relative velocity between target and missile. A low noise floor and a high resonator Q are desirable here, to allow weak targets to be detected at high relative velocities and also to minimize tracking errors in the terminal phase

As described, the power spectrum of Figure 1 yields little information about the coherence of the master

oscillator. Coherence is usually taken for granted, and quite properly so, when dealing with quartz-based resonators, or atomic resonance-stabilized clocks. As the oscillator quality, as defined by its resonator Q , deteriorates at microwave and MM-wave frequencies, and the width of the power spectral density about the carrier widens by orders of magnitude (see Figure 2), a quantitative measure of coherence becomes not only

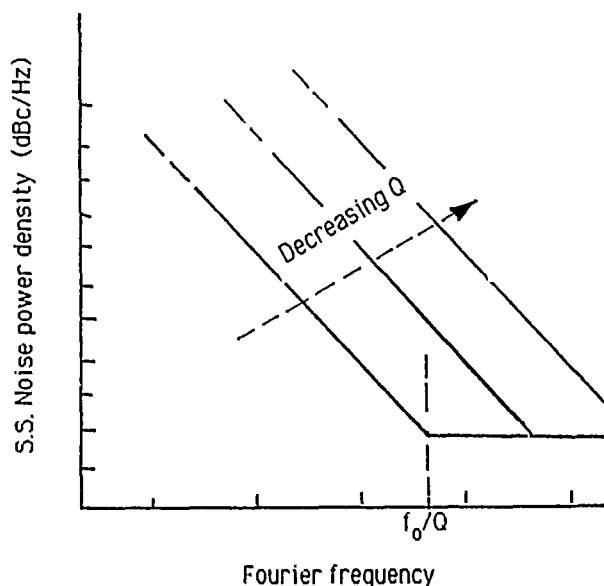


Figure 2 The effect of decreasing the resonator Q on spectral density and line width. The oscillator line width increases inversely with Q .

meaningful, but quite useful in special cases. The free-running MM-wave source as master oscillator in a map-making radar is a case in point.

Temporal Coherence

The significance of temporal coherence or frequency indeterminacy in the radar master oscillator is readily appreciated in the context of the mapping radar example. An estimate of the maximum radar range is required during which the free-running oscillator remains in step with the delayed radar return. Lack of adequate coherence results in degradation of the receiver sensitivity. In this paper we define a coherence interval τ_c during which the rms phase of the "noisy" oscillator deviates by one radian from a noiseless but otherwise identical source at the same frequency [3]. Additionally, we impose a safety margin of 10 times on the oscillator, so that the radar range round trip delay is limited to no more than $0.1 \tau_c$, which will be justified below. For the purpose of this treatment, slow phase changes due to thermal- or control voltage- induced frequency drift shall have been accommodated and are, therefore, not at issue in this discussion. Edson [3] defined a coherence

interval for thermionic sources, that is related to the resonator filter parameters and an electronic "noise temperature", given by

$$\tau_c = 2(P/kT_N)(Q/f)^2. \quad (1)$$

The first bracket contains the oscillator to thermal noise power ratio, where k is the Boltzman constant. The second bracket represents the resonator Q and frequency f . Note that this bracket has dimensions of time and represents the resonator's response time to a unit step function. Equation (1) suggests that the range of τ_c may vary over many orders of magnitude, as it amply demonstrated by the examples listed in Table I. A 1 MHz crystal oscillator with Q of 10^6 , operating at ambient temperature ($\approx 400^\circ \text{K}$) maintains coherence for 10^{14} seconds. A thermionic microwave oscillator, e.g. a reflex klystron at X-Band with a circuit Q of 10^3 and an electron temperature of $2 \times 10^4^\circ \text{K}$, may not remain coherent for more than 40 seconds by the criteria of equation (1). Unfortunately, the statistics of MM-wave solid state power sources are not readily characterized by a single parameter, such as noise temperature. For example, the

Table I

	Quartz	Klystron	FET/DRO
$f(\text{MHz})$	1	10^4	3×10^4
$P(\text{mw})$	1	10	10
Q	2×10^6	10^3	500
$T_N (^\circ \text{K})$	400	2×10^4	?
$\tau(\text{sec})$	$\approx 10^{16}$	≈ 40	?

noise temperature of a MM-wave FET source cannot be determined by noise temperature measurements of the amplifier using the same transistor. This is because the flicker ($1/f$) noise at lower frequencies within several hundred megahertz of the carrier raises the effective noise temperature substantially. Its direct relationship to the FET oscillator phase noise is under investigation [4]. The noise statistics of one-port sources, such as YIG-, Gunn-, and IMPATT oscillators are even less well defined. Given this dilemma, one is forced to search for an alternate means to estimate oscillator coherence.

Auto-correlation as a Measure of Coherence

One of the most meaningful measures of coherence is the auto-correlation function $R(\tau)$. Mathematically, $R(\tau)$ may be found from temporal as well as frequency-domain measurements via the Wiener-Kintchine theorem [5], as is shown in equation (2),

$$R(\tau) = \lim_{T \rightarrow \infty} (1/T) \int_0^T v(t) v(t+\tau) dt$$

$$= (1/2\pi) \int_{-\infty}^{+\infty} S(\omega) \exp(j\omega\tau) d\omega, \quad (2)$$

where $S(\omega)$ is the spectral power density of the noise-modulated carrier with amplitude $v(t)$. In the time domain, auto-correlation is performed by multiplication and integration of identical but time-delayed functions $v(t)$. The mixer and low pass (or IF) filter in the coherent radar receiver perform exactly this operation on delayed radar return and local oscillator signals.

Two examples suffice to show the useful frequency-time-domain relationship.

a) White noise through gaussian filter. The spectral density of white noise, passing through a gaussian low-pass filter is given by,

$$S(\omega) = (N/\sqrt{2\pi}\omega_0) \exp -\{(\omega/\omega_0)^2/2\} \quad (3)$$

and its corresponding auto-correlation function is

$$R(\tau) = (N/2\pi) \exp -\{(\omega_0\tau)^2/2\}. \quad (4)$$

Here N is the spectral noise power density per hertz and ω_0 is the radian frequency at which the gaussian noise power density is reduced to $\exp -0.5$ or 2.17 dB from its zero frequency value. A particular τ in equation (4), determined by $\omega_0\tau = 1$, as shown later in (7), might be a useful coherence measure of this hypothetical source.

b.) Inverse-square oscillator noise spectrum[6] The spectral phase noise density of a microwave oscillator is assumed to be

$$S_{\phi}(f) = 4\langle v^2 \rangle \tau_c / \{1 + (2\pi f \tau_c)^2\}, \quad (5)$$

with corresponding auto-correlation function, given by

$$R(t) = \langle v^2 \rangle \exp -(|t|/\tau_c), \quad (6)$$

where the brackets $\langle \rangle$ denote $R(0)$, the total normalized oscillator power. Equations (5) and (6) are shown in Figures 3 and 4. Note that $R(t=0.1\tau_c)$ has decreased to 0.9 according to our imposed coherence criteria.

Apparently, most fluctuating phenomena such as noise, give rise to this exponential form and a characteristic correlation interval τ_c [6] These two examples show that the noise distribution can be viewed in terms of the parameter $2\pi f_c$, where the one-sided noise spectrum width f_c , defined by $2\pi f_c = 1$, contains most of the

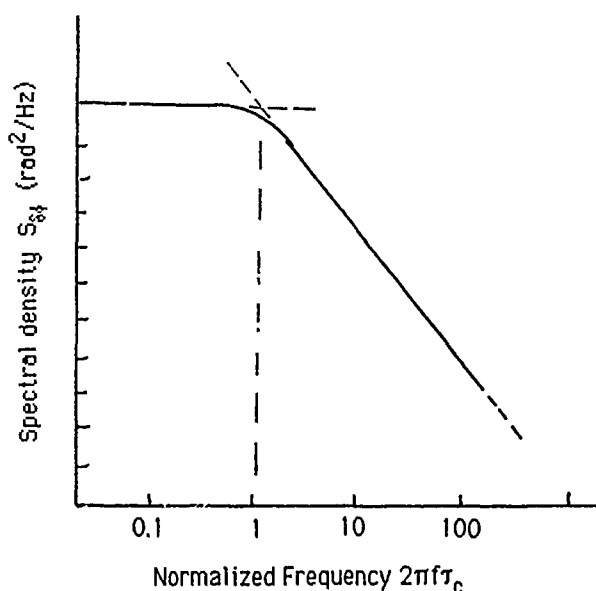


Figure 3. Spectral density function characterized by a single correlation interval τ_c .

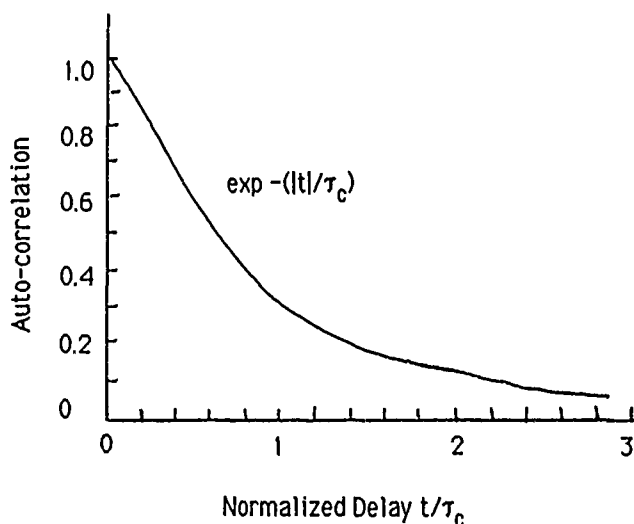


Figure 4 Auto-correlation of spectral density function shown in Figure 3.

carrier power. The two model distributions may be approximated by a constant noise spectral density region and a frequency-dependent region, intersecting at $2\pi f_c = 1$, as is shown in Figure 3. In what follows, we define, a priori a correlation interval as

$$\tau_c = 1/(2\pi f_c), \quad (7)$$

where f_c is the corresponding coherence frequency. Its significance will become apparent from the discussion to follow.

The Correlation Interval τ_c of an Oscillator

The noise performance of an oscillator is generally obtained in the frequency domain. Such a measurement yields the one-sided spectral density of noise power to carrier power ratio $L(f)$ (script L), where f , sometimes called the Fourier frequency, is measured from the carrier. The spectral density of rms phase fluctuations is obtained from $L(f)$, as shown earlier [1],

$$S_{\delta\phi}(f) = 2L(f) \text{ (rad}^2\text{/Hz)}, \quad (8)$$

where $\delta\phi$ is presumed to be a small angle. Put another way, $S_{\delta\phi}$ represents the rate at which the rms phase jitter varies in a given Fourier frequency range. Thus, the area under the $S_{\delta\phi}(f)$ versus f curve yields the rms phase difference between the noisy oscillator and the reference, presumed to be noise-free. Setting the area equal to 1 radian then defines f_c and in turn τ_c , using equation (7). We obtain the area from the definite integral of $S_{\delta\phi}$ between ∞ and f_c , such that the integral equals one radian, or

$$\int_{f_c}^{+\infty} S_{\delta\phi}(f) df = 2 \int_{f_c}^{+\infty} L(f) df = 1 \text{ (rad)}^2 \quad (9)$$

The shaded area in Figure 5 between f_c and ∞ represents one square radian. It is seen that the coherence interval

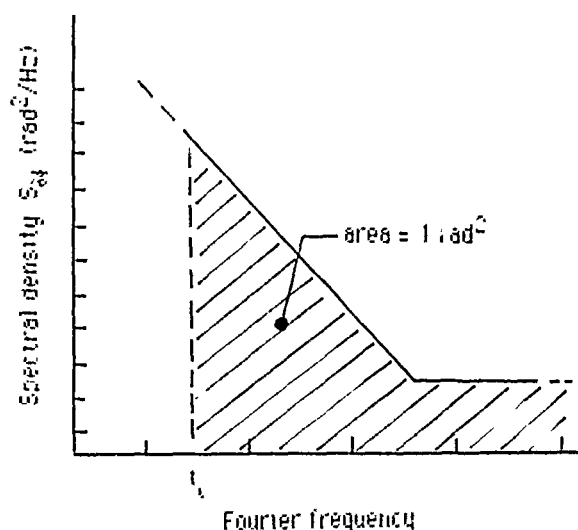


Figure 5 Cross-hatched area of integration between f_c and ∞ yields the oscillator line width f_c and correlation interval τ_c

τ_c may be obtained directly from the measured phase noise plot. In practice, little $S_{\delta\phi}$ contribution accrues from the higher Fourier frequencies near and beyond the noise floor. f_c also represents the minimum one-half

filter bandwidth through which most of the oscillator power can be passed and has been referred to as the "fast line width" of the oscillator [7]. We would expect, as in Figure 3, approximately constant spectral density at Fourier frequencies $f < f_c$. The simplified oscillator power spectrum of Figure 1 may now be modified to incorporate the additional corner frequency, as shown in Figure 6.

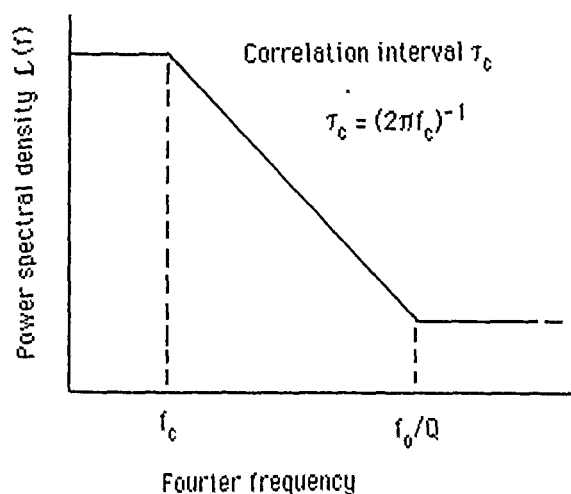


Figure 6 Simplified power spectral density of an oscillator with correlation interval τ_c and corner frequency f_c/Q .

A few examples from the published literature and from our own measurements suffice to illustrate the method. In Figure 7 the measured $L(f)$ for several representative microwave and MM-wave sources are shown, superimposed on the same plot. Their characteristics

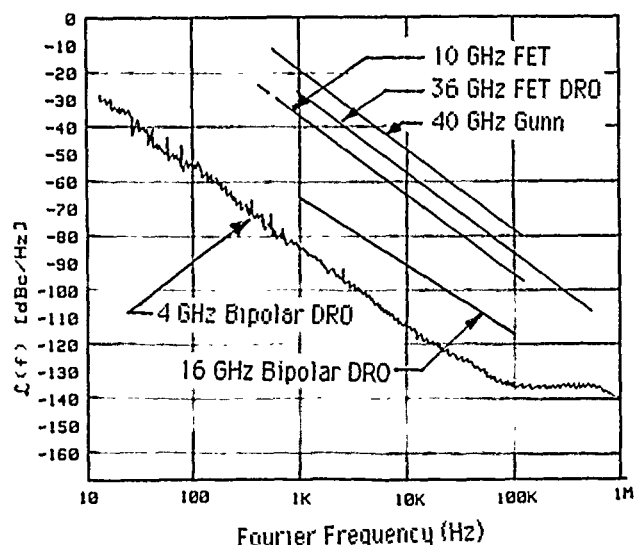


Figure 7 Measured power spectrum of several Microwave and MM-wave free-running oscillators

and computed coherence-related parameters, including the estimated maximum coherent radar range R , are listed in Table II.

Table II. Microwave & MM-wave Source Parameters.

Freq. (GHz)	Type	Power (dBm)	$L(10\text{KHz})$ (dBc/Hz)	f_c (KHz)	τ_c (μsec)	R (KM)	Ref.
4	Si-DRO*	+7	-114	2.8E-3	5.6E4	840	8
8.3	Si-DRO*	+10	-90	0.06	2.5E3	37.5	9
10	FET	+5**	-65	0.8	199	3.0	4
16	Si-DRO*	+7	-86	0.07	2.3E3	3.5	9
18	FETDRO	+10	-62	1.1	142	2.1	9
36	FETDRO	+10	-50	4.5	36	0.5	9
40	Gunn	+10**	-56	2.2	71	1.1	1

* - Bipolar, ** estimate.

It is noted that in the measured fourier frequency range, the $L(f)$ of these microwave sources exhibits f^{-3} dependence, a feature that permits the coherence interval to be simply computed if $L(f)$ is known in a narrow frequency interval. To do this we approximate the spectral phase noise density, as in (10),

$$S_{\phi} = 2 L(f) \approx a_3 f^{-3}, \quad (10)$$

which defines a_3 . Setting the integral in (9) equal to one square radian and using (7) through (10) we find for the coherence interval,

$$\tau_c \approx 1/(2\pi/2) \{ L(f_f) f_f^3 \}^{-1/2}, \quad (11)$$

where f_f is any narrow frequency band in which the measured $L(f)$ of the source is known. In deriving equation (11) it has been assumed that the frequency range over which $L(f)$ is integrated, is large. Thus, the portion of the integral that includes the noise floor contributes little to the accuracy of (11). The assumption is generally valid.

A corresponding relation may be found for $L(f)$ that exhibits other than f^{-3} dependence. In cases where the exponent varies in the fourier frequency range, the computation is carried out in several appropriate steps.

An Experiment to measure the Conerence Interval τ_c

An experimental approach is proposed to simulate the effects of long delay on receiver sensitivity when a "noisy" master oscillator is used. Schematically shown in Figure 8, the technique uses a phase-comparison bridge. One arm of the bridge constitutes the discretely variable radar range delay. It is simulated by a recirculating loop, using a 5- μsec broad band microwave

acoustic delay line in the 3 GHz frequency region. One half of the noisy S-Band source output is pulse modulated and feeds the delay arm, while the other arm

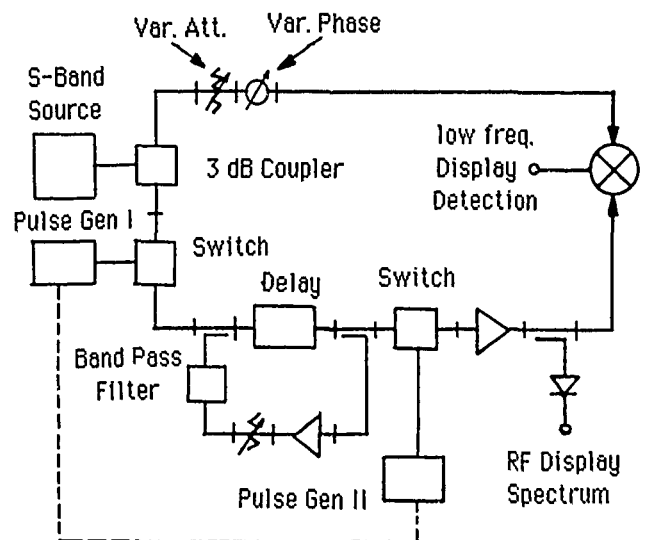


Figure 8. Proposed experiment for measuring the correlation interval using a recirculating loop.

simulates the CW local oscillator signal. Provisions have been made to coherently detect each of the delayed signals, after $5n$ μsecs , where $n = 1, 2, \dots$ are the number of round trips with adequate signal to noise ratio. The detected video amplitude decay versus n or delay will be used as a measure of coherence and to determine τ_c .

To date, the system, as shown in Figure 8, has been assembled and operated in a preliminary fashion. The spectrum of the 3.2 GHz pulse-modulated "noisy" source is shown in Figure 9. Video-detected pulses at the first three round trip delays ($\approx 5, 10$, and $15 \mu\text{secs}$) are shown in Figure 10. At this writing the experimental work is still in progress and results will be reported on completion at a later time.

Summary and Conclusions

Coherence of a "noisy" oscillator was shown to be a useful concept in certain radar applications. The coherence interval τ_c was defined as the time during which the rms phase of an oscillator deviates by one radian from a fluctuation-free reference. τ_c may be derived from the measured spectral density of phase noise. The autocorrelation function of a specific oscillator whose spectral density of phase noise exhibits f^{-2} fourier frequency dependence was used to relate τ_c to the source's line width f_c . Thus the oscillator noise spectrum may be approximated by two "corner"

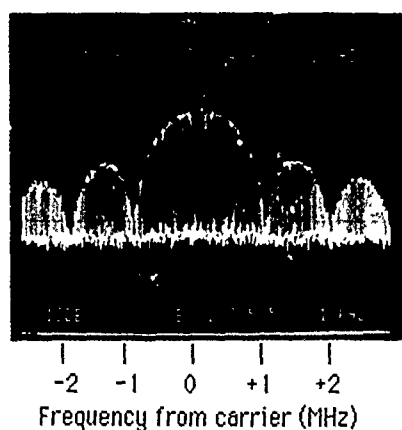


Figure 9. Power spectrum of the pulse-modulated "noisy" 3.2 GHz source.

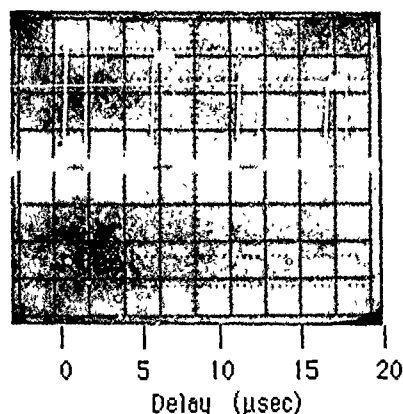


Figure 10. The time-domain showing the initial 1 μ sec video-detected pulse and three recirculated echos.

frequencies. It was suggested that microwave and MM-wave sources with frequency-dependent phase noise other than f^{-2} , may be similarly characterized. Further analytical and experimental work is required to support this premise.

The coherence interval of several sources between 4 and 40 GHz, all of which show approximately f^{-3} Fourier frequency dependence, was computed. An experimental technique was proposed by which the source coherence may be measured. Preliminary operational results were reported.

Acknowledgements

The author gratefully acknowledges fruitful discussions with D. Halford of the National Bureau of Standards, Boulder, CO., regarding the implications of the fast line width concept. Thanks also are extended to H. Leach of Hughes Missile Systems Group for his part in obtaining the experimental data.

References

1. A. L. Lance, W. D. Seal, F. G. Mendoza, and N. W. Hudson, "Automatic Phase Noise Measurements in the Frequency Domain", Proc. 31st Ann. Freq. Contr. Symp., pp 347-358, 1977.
2. D.W.Allan, "Statistics of Atomic Frequency Standards", Proc. IEEE, vol. 54, No. 2, pp 221-230, Feb. 1966.
3. W. A. Edson, "Noise in Oscillators", Proc. IRE, vol. 48, pp 1454-1466, Dec. 1960.
4. J. Graffeuil, K. Tantrarangroi and J. F. Sautereau, "Low Frequency Noise Physical Analysis for the Improvement of the Spectral Purity of GaAs FET's Oscillators", Solid State Electronics, vol. 25, No. 5, pp 367-374, ... 1982.
5. D. C. Champeney, Fourier transforms and their physical applications, (book), Apendix J, Academic Press, London & New York, 1973.
6. R. F. Voss, "1/f (flicker) noise: a brief review", Proc. 33th Ann. Freq. Contr. Symp., pp 40-46, 1979.
7. D. Halford, "Infrared-Microwave Frequency Synthesis Design: Some Relevant Conceptual Noise Aspects", Proc. Frequency Standards and Metrology Seminar, Quebec, Can. vol. 30, pp. 431-466, August 31, 1971.
8. kindly supplied by C. E. Samuelson, Hughes Aircraft Co.
9. kindly supplied by R. Oyafuso, Avantek.

THE GAAS FET OSCILLATOR-ITS SIGNAL AND NOISE PERFORMANCE

Robert A. Pucel
Raytheon Co.
131 Spring Street
Lexington, MA 02173

Abstract

A general review will be presented of the operating principles of the GaAs FET as a nonlinear device, with special emphasis on its properties as an oscillator element. The performance characteristics as an oscillator will be emphasized, with special attention paid to the mechanism by which the near-carrier noise is generated.

Introduction

The performance of the voltage-controlled oscillator (VCO) is critical to the success of many microwave systems, especially radar systems. Until recently, the only viable solid-state sources were the bipolar transistor at the lower microwave frequencies (L- to C-bands) and the Gunn and IMPATT(avalanche) diodes at the higher frequencies.

During the last ten years the Schottky-gate gallium arsenide field-effect transistor (GaAs MESFET) has demonstrated its superiority as a low-noise amplifying element. The intrinsic noise mechanisms responsible for this good noise performance are well-understood.

More recently the GaAs FET has demonstrated its unique advantages as an easily-designed, wide-band, and high frequency oscillator component. The GaAs FET oscillator permits direct generation of microwave signals with good efficiency and temperature stability. The FET's submicron geometry offers the potential for operation in the millimeter range. Tuning of the oscillator's frequency with YIG devices or varactors increases the utility of this device.

For example, octave-tunable FET oscillators are common place. Oscillator operation in the laboratory has been demonstrated in the millimeter range to frequencies in excess of 100 GHz. FET oscillators have been incorporated into microwave monolithic circuits on a GaAs chip (MMICs).

The GaAs FET has one major drawback which makes it unacceptable for many systems applications, namely, its rather high level of near-carrier (1/f) noise, especially in the phase spectrum. This noise is generated at baseband and is up-converted to the carrier band through the (nonlinear) mixing mechanisms associated with oscillator operation.

In all cases where oscillator noise spectra have been reported, the near-carrier noise level of GaAs FETs has been higher than that of bipolar transistors and Gunn and avalanche diodes, unless extreme measures are invoked, such as the use of temperature-compensated high-K dielectric cavities and other methods of frequency stabilization.

Scope of Presentation

This paper will begin with a general review of the operating principles of the FET as a nonlinear device with special emphasis on its use as an oscillator. The relation between its performance as a power amplifier and an oscillator will be illustrated. The performance characteristics as an oscillator, both signal and noise will be emphasized.

A survey will be presented listing the best power-efficiency and noise properties reported by various laboratories for the frequency range spanning from C-band to 100 GHz. The wide-band tuning capabilities which have been achieved to date over the microwave and millimeter bands will be discussed.

Special attention will be focussed on the near-carrier noise generation in GaAs FET oscillators. The mechanisms responsible for this noise generation will be described, and a model will be presented to explain the means by which this near-carrier noise arises by frequency upconversion of baseband noise. Experimental results obtained by the author verifying the model will be presented.

Finally, the various means for stabilizing GaAs FET oscillators to reduce this near-carrier noise will be reviewed. Both discrete (hybrid circuit) and monolithic circuit oscillators will be considered.

Review of FET Principles

A cross section of a gallium arsenide (GaAs) Field-Effect Transistor (FET) is illustrated in Fig. 1. Shown is a body of semiconductor with three electrodes at the surface. The two electrodes labelled S and D are ohmic contacts representing the source (of electrons) and the drain, respectively. The conducting path represented by the semiconductor (GaAs) constitutes the source-drain or output circuit. The source is usually held at ground potential, though this is not a necessity. The drain, or output terminal is biased positively with respect to the source.

The third electrode, labelled G, denotes the gate contact. Unlike the source and drain contacts, it is not ohmic but rather a metal-semiconductor junction. Under normal operation, this junction is reverse-biased with respect to the source (gate at a negative bias potential). Being reverse biased, the gate-source path conducts a negligible current.

The barrier potential at the gate-semiconductor interface, re-enforced by the negative bias voltage applied externally to the gate (labelled V_{gg} in Fig. 1) produces a negative charge on the gate which must be balanced by a corresponding positive charge in the underlying semiconductor. This positive charge is supplied by the donor (dopant) atoms which have been "stripped" of their free electron by the reverse bias voltage on the gate. These removed electrons flow to the source electrode. The result is that a portion of the semiconductor under the gate is devoid of conducting electrons, and behaves, as a consequence, as an insulator, albeit a charged insulator. The extent of this region, denoted as the depletion zone, is a function of the applied gate-source bias voltage. The larger in magnitude this bias, the greater in extent is the depletion zone. The depletion zone is indicated by the shaded region in Fig. 1.

The depletion zone, because it is devoid of electrons, constrains the source-drain current to a small cross section. In effect, the gate voltage controls the size of the conducting

cross section, hence the magnitude of source-drain current. The portion of semiconductor under the gate participating in conduction is often called the channel. It follows that if a small signal voltage is super-imposed on the gate bias, the channel cross section will be modulated, as will be the channel or output current. This is the mechanism by which the FET can exhibit power gain, since little input power is required to produce a much greater signal power in the output circuit. As we shall see, this channel modulation process is also the mechanism by which baseband noise is translated into the carrier band of a FET oscillator.

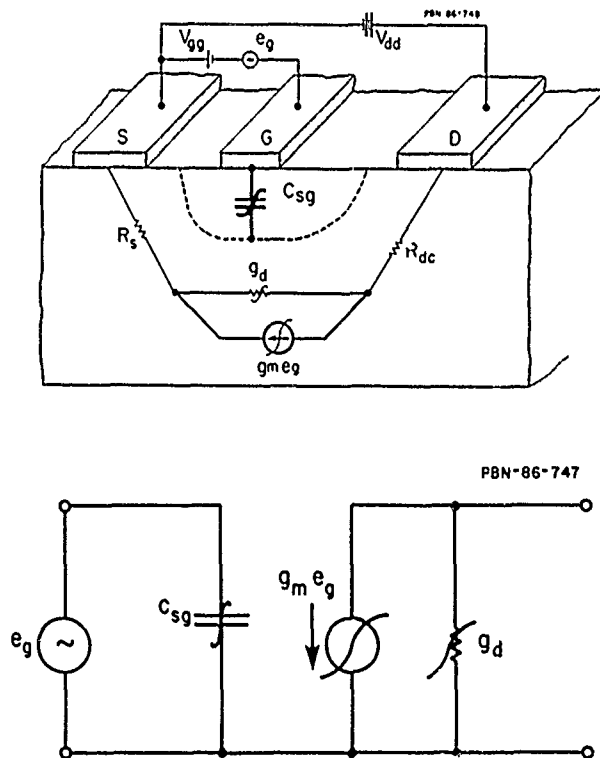


Fig. 1 Nonlinear Model of an FET

Equivalent Circuit of the FET

It can be shown by circuit modelling that the depletion layer of an FET behaves as a nonlinear capacitor, that is, a capacitor which is a nonlinear function of the applied gate-source bias. This capacitance is illustrated in the simplified equivalent circuit of the FET shown in the bottom of Fig. 1.

The current modulation mechanism described above is represented by the voltage-controlled current generator labelled $g_m e_g$, where e_g represents the signal voltage applied to the input (gate-source terminals). The modulation process is denoted by the transfer "function" g_m , called mutual conductance or transconductance. In general, g_m is a nonlinear function of the gate source bias voltage. The source-drain channel or output conductance is represented by the symbol g_d . This output conductance is a nonlinear function of V_{gg} and the source-drain bias voltage V_{dd} .

The FET as a Large-Signal Device

The FET acts as a linear transducer if the input power level is low enough as shown in the top-most graph in Fig. 2. However, as the signal level is increased, such that

g_m and g_d become signal level dependent, the power gain characteristic, denoted by the upper branch (labelled P_{out} in the graph), begins to "saturate", as is typical of most amplifying devices.

Consider, now, an oscillator constructed from this amplifier by connection of a suitable feedback path from output to input as shown in the lower sketch in Fig. 2. For convenience, we assume the feedback circuit to be lossless and to have the proper phase shift to sustain oscillations. From the power gain characteristic of the amplifier, it is easy to convince oneself that the power output characteristic of the corresponding oscillator (denoted by the branch labelled $P_{out}-P_{in}$ in Fig. 2), will exhibit a peak at a certain input power level (supplied by feedback). The incremental gain of the corresponding amplifier is unity at this operating condition. FET oscillators are usually operated at this optimum condition for maximum power output.

It is obvious that the maximum available power output of an FET oscillator cannot exceed that of the corresponding amplifier. Note that the optimum operation occurs when the FET is operating only slightly out of the linear regime. Because of this, FET oscillators are especially easy to design, since linear techniques can be used to a great extent to obtain the initial design.

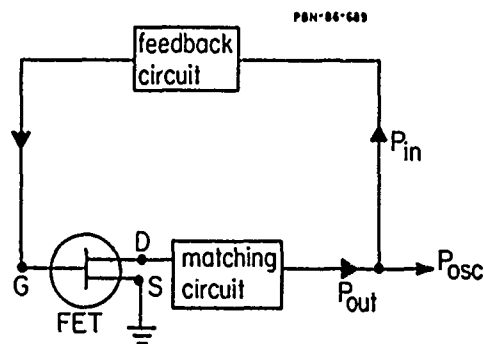
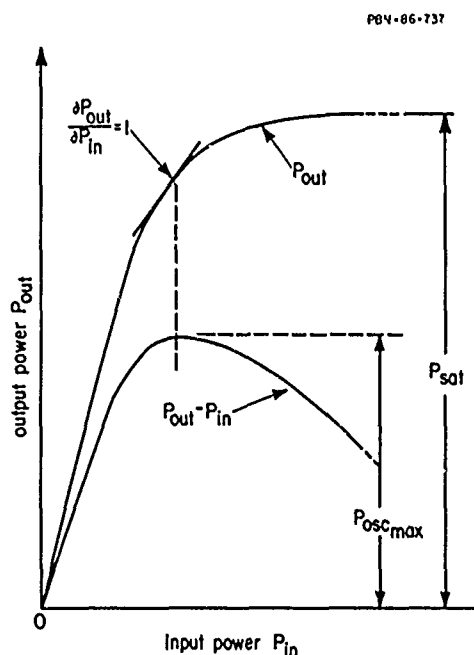


Fig. 2 Power Characteristic of an FET Amplifier and Oscillator

The Power Performance of the FET Oscillator

The FET, as an oscillator, is easily tuned by an inductive element at one of its ports, say the gate-source port, as shown in Fig. 3. Illustrated is an inductive element, in conjunction with a varactor diode. The combination presents a voltage controlled inductive element to the FET which will resonate the gate-source capacitance. Other methods of tuning are possible, as we shall see.

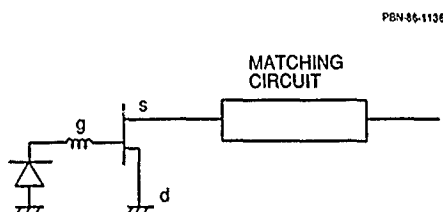


Fig. 3 Narrow-Band Tuning of an FET Oscillator

With simple varactor tuning of the type shown, oscillator frequencies can be varied over a 30% band with little difficulty. FET oscillators also can be resonated by other means such as dielectric resonators. This type of FET oscillator is labelled DRO. DROs, of course, can only be tuned over a few percent band, if that.

Figure 4 summarizes the best oscillator performance obtained with GaAs FETs reported by various laboratories for narrow-band operation such as obtained with single-varactor voltage tuning (VCO) or dielectric resonators. The names indicated in the graph represent the laboratory at which the performance was obtained. Also indicated is the DC-to-RF efficiency.

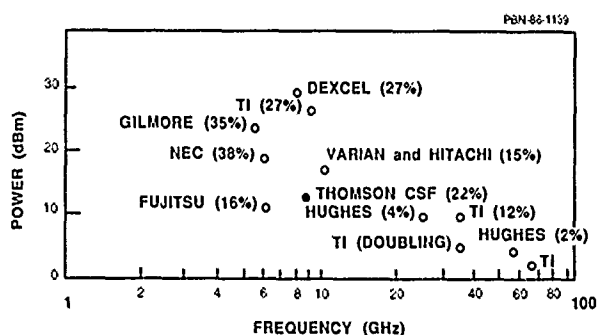


Fig. 4 Survey of Narrow-Band FET Oscillator Performance

Most of these results were obtained in hybrid integrated circuits. Recently the upper frequency limit has been extended to 115 GHz with a monolithic circuit as reported by Tserng.² The power output for this case was small, namely, 0.1 mW.

Perhaps more interesting is the wide-band performance of FET oscillators. Since the FET is intrinsically a broadband device, it can be tuned over a multi-octave band

by appropriately designed tuning circuits. These circuits can employ several varactor elements, for example, as shown in Fig. 5, or YIG spheres. The result is the remarkable broadband performance illustrated in Fig. 6. Note that tuning ranges exceeding two octaves have been achieved.

We turn, next, to the main topic of interest, the noise performance of FET oscillators.

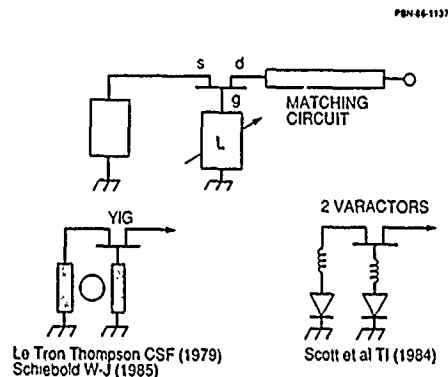


Fig. 5 Wide-Band Tuning of an FET Oscillator

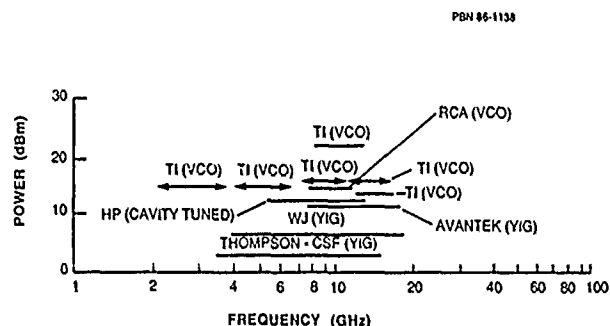


Fig. 6 Survey of Wide-Band FET Oscillator Performance

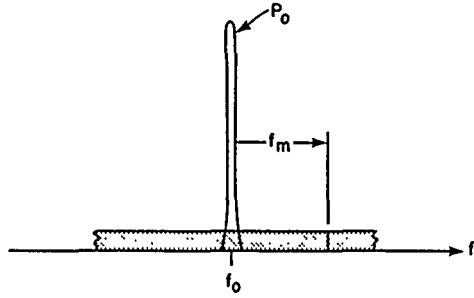
Sources of Noise in FET Oscillators

Noise in an FET oscillator manifests itself in two forms as illustrated in Fig. 7. In the top figure is illustrated a white background source in the carrier band. This noise arises from thermal and shot-noise generated in the source-drain channel within the operating frequency band of the oscillator. It is not a significant contribution to FET oscillators, being important only at offset frequencies far from the carrier. However, this source of noise is the predominant noise mechanism for FET amplifiers.

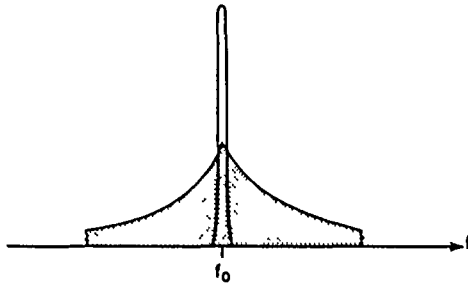
The most important noise in an FET oscillator is the so-called "near-carrier" spectrum which follows, closely, the $1/f$ spectral distribution sketched in Fig. 7(b). This noise arises from a baseband source which is upconverted to the carrier band by a modulation process to be described below.

Baseband noise originates from random fluctuations in the source-drain channel current produced by the random capture and release of electrons by "traps" in the semiconductor body. The trapping mechanism is especially pronounced in GaAs material. The trapping centers (traps) can arise from a variety of causes such as trace impurities, crystal defects, and surface states attributable to contaminants and other causes.³

The effect of these traps, namely, fluctuations in channel current, must be accommodated by fluctuations in the



(a) Background Noise



(b) Up-Converted Baseband Noise

Fig. 7 Noise Spectral Components of an FET Oscillator

channel cross section, that is by fluctuations of the depletion layer boundary. Since depletion layer fluctuations are equivalent to modulation of the gate-source capacitance, it follows that under oscillation conditions, the baseband current fluctuations manifest themselves as a carrier frequency modulation. For this reason, this type of noise is often called modulation noise. It should be stated that modulation noise also manifests itself as carrier amplitude fluctuations (AM) noise, but the modulation process is too weak to make this noise contribution important. In summary, modulation noise is manifested as a frequency modulation (FM) spectrum about the carrier as illustrated in Fig. 7(b).

The noise upconversion process can be represented very simply in mathematical terms. Let the mean-square fluctuation in baseband current be represented by $\langle (\Delta i_d)^2 \rangle$, where $\langle \rangle$ denotes a statistical average. This fluctuation is "equivalent" to a random fluctuation of the gate voltage ΔV_g whose mean square value is given by

$$\langle (\Delta V_g)^2 \rangle = \frac{\langle (\Delta i_d)^2 \rangle}{g_m^2} \quad (1)$$

(See Fig. 1.) Here g_m denotes the transconductance averaged over an oscillation cycle.^{3,4}

Let the sensitivity of the oscillation frequency to a small change in gate voltage be denoted by a sensitivity factor S given by

$$S = \frac{\partial f_0}{\partial V_{gg}} \quad (2)$$

where f_0 represents the oscillator frequency. It follows that the mean-square fluctuation in the carrier frequency is given

by the expression

$$\Delta f_{rms}^2 = \langle (\Delta f_0)^2 \rangle = |S|^2 \langle (\Delta V_g)^2 \rangle \quad (3)$$

$$= \frac{|S|^2}{g_m^2} \langle (\Delta i_d)^2 \rangle \quad (4)$$

where the baseband noise $\langle (\Delta i_d)^2 \rangle$ is measured under oscillating conditions.

Experimental verification of this model is illustrated by the results depicted in Fig. 8. Shown are three graphs. The lowest graph represents the measured spectrum of the baseband noise as a function of baseband or modulation frequency f_m . The topmost graph is the predicted FM noise level as a function of frequency offset from the carrier (10 GHz), also denoted by f_m . The intermediate graph is the measured FM noise spectrum. Note that the measured and predicted FM spectra differ by only 1 dB!

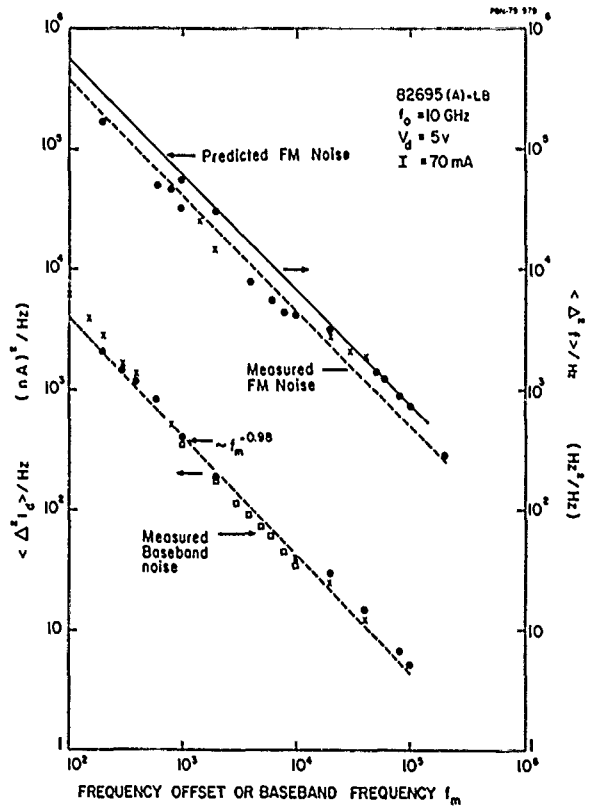


Fig. 8 Relation of Baseband Noise to FM Noise

What noise performance has been reported to date? Figure 9 is a survey graph listing results reported in the literature for an offset frequency of 10 kHz from the carrier for a range of oscillation frequencies. (Note that, theoretically, the FM noise spectrum should be independent of carrier frequency.) The vertical scale refers to the noise level in a Hz bandwidth referenced to the carrier power. The nomenclature in parentheses denotes the method of frequency stabilization or tuning used, i.e. VCO, DRO, etc. The isolated result by Raytheon denoted by a vertical arrow refers to a DRO with noise degeneration. This method,

which has achieved the lowest reported noise-to-carrier ratio with FETs, will be described later.

PEN 86-1140

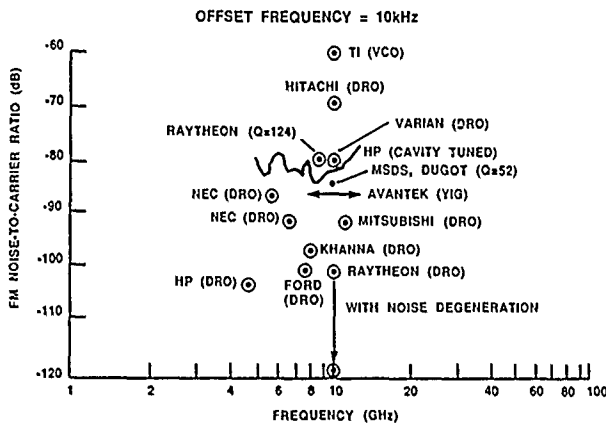


Fig. 9 Survey of FET Oscillator Noise Performance

Frequency stabilization of FET oscillators in hybrid circuits usually is accomplished with dielectric resonators coupled to microstrip stub sections. Figure 10 illustrates the variety of circuit formats that have been employed with dielectric resonators.⁶⁻¹⁰ (The shaded elements represent microstrip stubs.)

PEN 86-1141

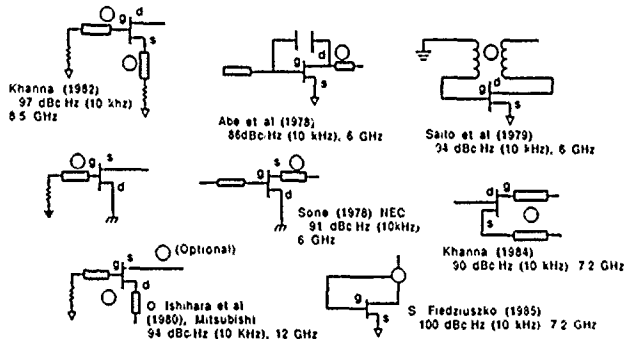


Fig. 10 Methods of Frequency Stabilization

A clever approach reported by Galani et al, illustrated in Fig. 11, uses the dielectric resonator as the frequency stabilizing element in the usual manner; however, in addition, it also capitalizes on the frequency dispersive properties of the resonator as a reflective element in a frequency-locked loop to produce noise degeneration.¹⁰ This technique employs a dielectric resonator in a feedback loop to set the frequency of oscillation (the upper feedback path in Fig. 11.) Part of the power directed to the feedback path is reflected by the dielectric resonator into one port of a phase detector where it is compared to a reference signal entering the second port of the phase detector from the oscillator output. By a suitable phase shift of this reference signal, the output of the phase detector represents the FM noise of the oscillator translated to base band. The baseband noise is amplified and sent back to the gate terminal of the oscillating FET as a negative feedback signal which partially cancels the original baseband modulation noise. The improvement in performance with noise degeneration is demonstrated in Fig. 12. To date, this method has achieved the lowest N/C ratio obtained with FET oscillators.

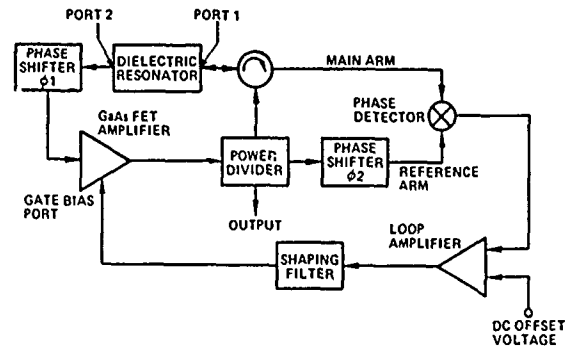


Fig. 11 Circuit for Noise Degeneration in FET Oscillators

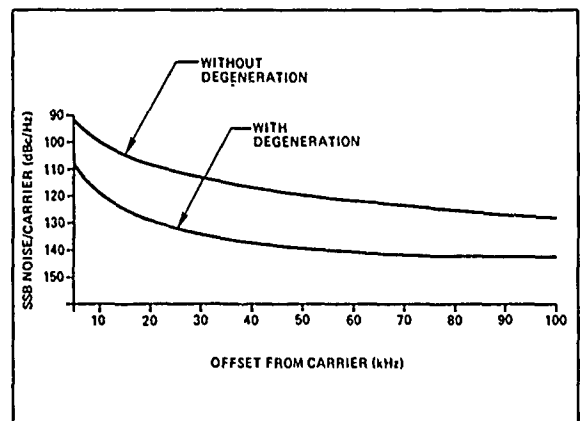


Fig. 12 Noise Performance of an FET Oscillator With and Without Noise Degeneration

Comparison of the Noise Performance of FET Oscillators with Other Solid-State Sources

It is informative to compare the inherent noise properties of FET oscillators with that exhibited by other solid-state devices such as bipolar transistors, Gunn diodes, and IMPATT diodes. In order to provide a fair comparison, independent of the circuit Q factor or method of stabilization, we must compare the intrinsic noise performance of these devices. For this purpose we express the mean square frequency deviation of an oscillator in the form¹¹

$$\langle (\Delta f)^2 \rangle = \left(\frac{f_o}{Q_L} \right)^2 \frac{M(f_m)kTB}{P_o} \quad (5)$$

where $Q_L' = Q_e Q_L / (Q_e + Q_L)$, and Q_e and Q_L are the external and loaded Q-factors, respectively. Here kTB is the reference thermal noise power in a bandwidth B at a temperature T , and f_o and P_o represent the oscillator frequency and carrier power output, respectively.

The factor M , the noise measure, embodies all of the intrinsic noise properties of the solid state device. The noise measure, though originally applied to the amplifying mode¹², and specifically to the white noise background spectrum, was extended to oscillators by Haus, et al¹³, and to upconverted noise by Kurokawa.¹¹ It is a property of the device, alone. In effect, M is the factor by which the available noise power

of the device in the oscillation mode exceeds the thermal reference level kTB . In the limit of small-signal oscillations, M corresponds to $F-1$ where F is the noise figure of the corresponding amplifying device operating at the same bias voltages. Thus, a comparison of the noise measures of the various solid state devices is a valid means of rating their relative noise performance when operated as oscillators.

Noise measure is most conveniently plotted as a function of the frequency offset from the carrier as illustrated in Fig. 13. Shown are four (simplified) graphs which provide a semi-quantitative comparison between the four major solid-state microwave devices. These curves should be considered guides rather than design tools.

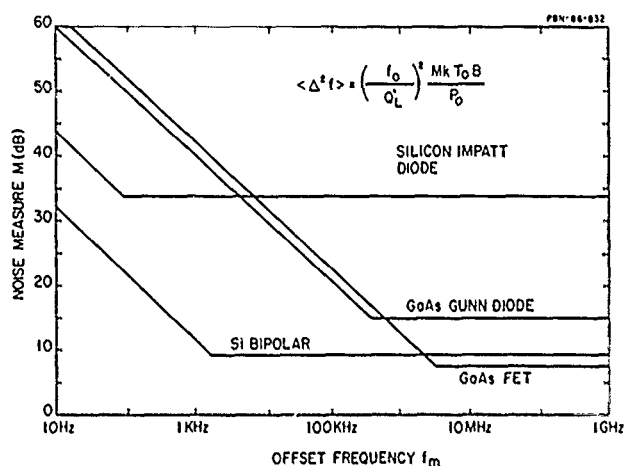


Fig. 13 Comparison of Noise Measure of Four Types of Solid State Oscillators

The first element of note is that the spectrum of M consists of a near-carrier segment decreasing at 10 db/decade ($1/f$) merging into a flat (white noise) segment. The $1/f$ spectrum, of course, arises from the modulation noise, the flat portion, from the background noise.

Observe that the bipolar transistor exhibits the best (lowest) near-carrier noise performance, followed by the silicon IMPATT diode, then the GaAs Gunn diode, and finally the FET. It is not a coincidence that the noisiest two devices are made from GaAs. It is well-known that GaAs devices have a higher baseband noise than silicon devices. On the other hand, in the flat spectral zone, the FET is the least noisy, followed by the bipolar transistor, then the Gunn device, and finally the IMPATT. Actually, to be fair to the Gunn diode, somewhat better performance is obtained with InP Gunn devices. The low white-noise spectrum of the FET signifies the superior noise performance of this device as an amplifier, since the same noise spectrum is responsible for the low noise figure in this mode of operation.

Lest one conclude that the bipolar transistor is the device of choice for oscillator applications, one should remember that it is the most limited in frequency range of all four types of devices. The bipolar, though it has been demonstrated to operate as high as 20 GHz in the laboratory, is, for all practical purposes, relegated to the frequency range from C-band and down. The FET, Gunn, and IMPATT oscillators, on the other hand, have been operated up to and beyond 100 GHz. Thus, which device should be favored for a given application will depend on the oscillation

frequency range as well as the noise performance. The bipolar transistor appears to be the unquestioned choice for frequencies below, approximately, 4 GHz.

Summary and Conclusions

We have attempted to give the reader an appreciation of the capabilities and limitation of the FET oscillator in this paper. It has been pointed out that the FET oscillator is especially easy to design because, in the optimum mode of operation (maximum power), it is only slightly in the non-linear regime, thus linear design techniques can be utilized to a great extent.

We have demonstrated that the FET oscillator is inherently broadband, capable of being tuned over several octaves, and can be operated well into the millimeter band if designed properly.

It was shown that the near-carrier noise is predominantly FM noise, and is higher than competing solid-state devices. The principle source of this noise is upconverted baseband noise attributable to traps. To minimize this noise, one must operate FET oscillators with high-Q circuits, such as provided by dielectric resonators. Unfortunately, since the basic trap mechanisms generating the baseband noise are not well-understood, the reduction of this intrinsic noise by any significant amount does not seem likely in the foreseeable future.

Acknowledgement

I take pleasure in acknowledging the help of E. Niehenke of Westinghouse, who supplied me with much of the survey material used in this paper. I also am grateful to T. Parker and G. Montress of Raytheon with whom I had useful and informative discussions on alternative ways of looking at noise analysis of oscillators.

References

1. R. A. Pucel, R. A. Haus, and H. Statz, "Signal and Noise Properties of Gallium Arsenide Microwave Field-Effect Transistors," *Advances in Electronics and Electron Physics*, Vol. 38, Academic Press, New York (1975).
2. H. Q. Tserng, "A 115 GHz Monolithic GaAs FET Oscillator," *GaAs IC Symposium Technical Digest*, pp. 11-12 (1985).
3. R. A. Pucel, FET Noise Studies, Final Technical Report, Contr. No. F49620-79-C-0024, AFOSR, March 1981.
4. R. A. Pucel, FET Noise Studies, Final Technical Report, Contr. No. F49620-C-0052, AFOSR, Oct. 1982.
5. H. J. Siweris and B. Schiek, "Analysis of Noise Upconversion in Microwave FET Oscillators," *Trans. on Microwave Theory and Techniques*, Vol. MTT-33, pp. 233-242, March 1985.
6. A. P. Khanna, "Parallel Feedback FETDRO Design Using 3-Port S-Parameters," 1985 IEEE MTT-S International Microwave Symposium Digest, pp. 181-183.
7. T. Saito, Y. Arai, H. Komizo, Y. Itoh, and T. Nishikawa, "A 6 GHz Highly Stabilized GaAs FET Oscillator Using A Dielectric Resonator," 1979 IEEE MTT-S International Microwave Symposium Digest, pp. 197-199.

7. A. P. Khanna, J. Obregon, and Y. Garault, "Efficient Low-Noise Three Port X-Band FET Oscillator Using Two Dielectric Resonators," 1982 IEEE MTT-S International Microwave Symposium Digest, pp. 277-279.
8. H. Abe, Y. Takayama, A. Higashisaka, and H. Takamizawa, "A Highly Stabilized Low-Noise GaAs FET Integrated Oscillator with a Dielectric Resonator in the C Band," IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-26, pp. 156-162, Mar. 1978.
9. G. Ishihara, T. Mori, H. Sawano, and M. Nakatani, "A Highly Stabilized GaAs FET Oscillator Using a Dielectric Resonator Feedback Circuit in 9-14 GHz," IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-28, pp. 817-824, Aug. 1980.
10. Z. Galani, M. Bianchini, R. Waterman, R. Dibiase, R. Laton, and J. Cole, "Analysis and Design of a Single-Resonator GaAs FET oscillator with Noise Degeneration," IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-32, pp. 1556-1565, Dec. 1984.
11. K. Kurokawa, "Noise in Synchronized Oscillators," IEEE Trans. on Microwave theory and Techniques, Vol. MTT-16, pp. 234-240, April 1968.
12. H. A. Haus and R. B. Adler, Circuit Theory of Linear Noisy Networks, Technology Press of Mass. Inst. of Technology, Cambridge, MA (1959)
13. H. A. Haus, H. Statz, and R. A. Pucel, "Optimum Noise Measure of IMPATT Diodes," IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-19, pp. 801-813, Oct. 1971.

MAGNETICALLY TUNABLE HIGH OVERTONE MICROWAVE RESONATORS

J. D. Adam and S. R. McAvoy
Westinghouse R&D Center
Pittsburgh, PA 15235

H. L. Salvo, Jr.
Westinghouse Defense & Electric Center
Baltimore, MD 21203

Abstract

The development of yttrium iron garnet (YIG) resonators excited to overtone resonance by thin film ZnO transducers is described. Acoustic waves propagating in the YIG can couple by magnetoelastic interaction to spin waves and hence the resonance can be changed in frequency by an externally applied magnetic field. Single crystal YIG has low shear mode acoustic loss and is the lowest loss ferrimagnetic material known. Results are described for both shear mode overtones along the $\langle 1,1,1 \rangle$ direction and longitudinal mode overtones along the $\langle 1,1,0 \rangle$ path. The applications for the resonators described include tunable microwave filters and tracking oscillators.

Introduction

The high overtone bulk acoustic resonator (HBAR) has been shown to provide loaded Q's in excess of 50,000 at 1.5 GHz.¹ Temperature stable operation has been demonstrated using bulk shear mode resonances in selected cuts of lithium tantalate.² Low noise STALO's with frequency agility have been fabricated using HBARs.¹ Additional applications would be afforded the HBAR if a relatively simple means of continuously tuning the resonator were available. One such application is shown schematically in Figure 1. As indicated in the figure the HBAR is thermally tuned by means of a feedback circuit to a heater which is an integral part of the resonator mount. In this configuration yttrium aluminum garnet, YAG, $\langle 1,1,1 \rangle$, has been used as the resonator substrate exhibiting a temperature coefficient of -48 parts per million per degree centigrade. The mass of the resonator and mounting is such that the thermal tuning time constant is several seconds. It is apparently not feasible to reduce this mass to the point that the tracking response is fast enough for all applications. An alternate means of tuning is therefore very desirable.

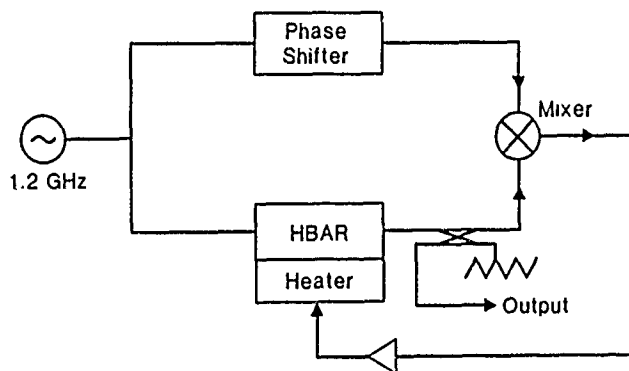


Figure 1. Thermally tuned YAG HBAR (High Overtone Bulk Acoustic Resonator) in a tracking filter configuration.

Magnetically tuned dielectric resonators have been available since the late 1970s.³ As an example, a strontium titanate disk is bonded to a YIG disk of the same diameter. Tuning is achieved by mode coupling between the electromagnetic resonances of the dielectric resonator and the ferromagnetic resonance of the YIG. At 9 GHz a 10% tuning range is obtained for an applied field change of about 1.0 kG. One potential drawback of this approach is the large magnitude and change in the applied magnetic field required with the attendant large power requirement. The principal advantage is that filters constructed of resonators of this type are usable at power levels of up to one watt without limiting. It is desirable on balance, to explore other possibilities with the goals of small size and weight and low power consumption.

Tuned YIG HBAR

A cross-sectional schematic view of one configuration of the magnetically tuned resonator is shown in Figure 2. The arrow in the center shows the direction of the body diagonal $\langle 1,1,1 \rangle$ of the YIG which is also the direction of the acoustic path and hence of the standing wave. This wave is generated and detected by a pair of thin film ZnO transducers disposed as shown in the figure. The direction of the externally applied magnetic field is shown by the arrow to the left of the resonator. The standing acoustic wave can couple by magnetoelastic interaction to spin waves.⁴ Strong interactions occur only close to the cross-over of the dispersion curves for acoustic and spin waves. The strength of the interaction, i.e., the branch splitting at cross-over, depends on the magnetoelastic coupling and the crystal orientations.

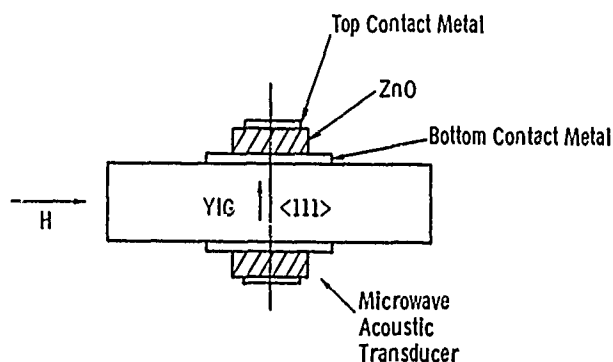


Figure 2. Cross-sectional schematic view of YIG HBAR showing placement of ZnO transducers and the directions of acoustic propagation and external magnetic field (H).

Experimental Results

Acoustic Q values in excess of 10^5 have been observed in YIG using bulk shear waves at 1.0 GHz. Shear waves for the current series of experiments are generated using transducers with thin film ZnO which is fabricated with the c -axis tilted with respect to the substrate normal. This approach, using an optimized transducer design for best coupling, provides the maximum loaded Q available acoustically. Untuned, loaded Q values in the range of 15×10^3 to 30×10^3 at 1.6 GHz are measured. Work currently is being performed on the optimization of resonator loaded Q . A computer model is being used to determine the optimum transducer dimensions for the trade-off between such parameters as diffraction effects and external circuit loading. The effects of magnetic losses on the resonator Q are also under evaluation.

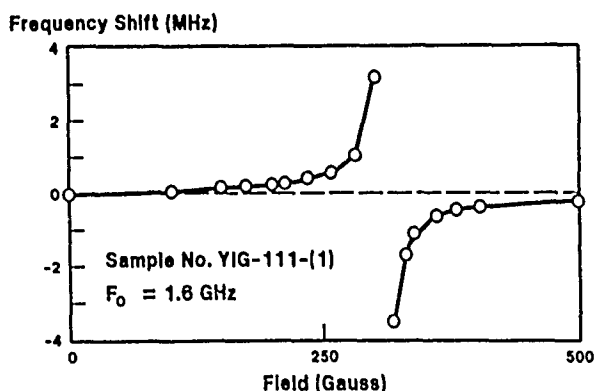


Figure 3. Tuning curve obtained for the resonator described in Figure 2. Frequency of resonator overtone is near 1.6 GHz.

Figure 3 shows the tuning data obtained for a resonator depicted in Figure 2, i.e., a shear mode HBAR excited along the YIG body diagonal with the applied magnetic field in the plane of the YIG plate. For magnitudes of applied field just above 250 gauss the frequency shift is seen to approach 4 MHz. Typically these resonators have overtone spacing of 5 MHz so the observed frequency shift is adequate to tune between overtones.

The temperature coefficient of frequency of these resonators has been measured to be -60 ppm/ $^{\circ}\text{C}$ for longitudinal mode resonances propagating along $\langle 110 \rangle$. A demonstration of frequency stabilization by means of the external field was performed using such a resonator. In this case longitudinal mode ZnO film transducers were used to excite the resonator. The results are shown in Figure 4. The magnetic field is changed by about 50 gauss to hold the frequency constant at 1.604 GHz while the temperature is changed from 23°C to 33°C .

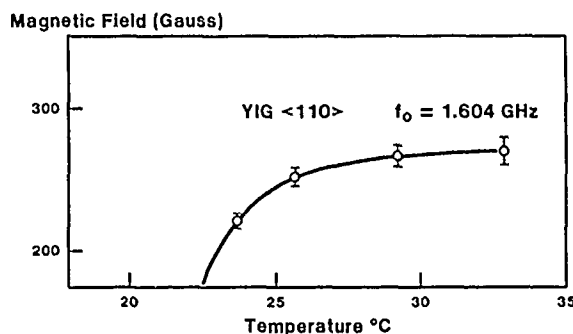


Figure 4. Magnitude of the external field required to hold the frequency constant at 1.604 GHz as the temperature is changed.

Conclusions

Frequency tuning has been demonstrated in shear mode HBARS using YIG with a $\langle 1,1,1 \rangle$ propagation direction and for longitudinal mode resonances along the $\langle 1,1,0 \rangle$ direction. Although shear waves provide lower propagation losses and hence higher Q resonators, longitudinal mode transducers are more easily produced and at present offer lower insertion loss devices. It is believed that application opportunities exist for both in tunable filters and tracking microwave oscillators.

References

1. J. T. Haynes, et al., Stable Microwave Source Using High Overtone Bulk Resonators," 1985 IEEE MTT-S Digest, pp. 243-246.
2. B. R. McAvoy, et al, "Highly Stable, Ovenized Bulk Shear Mode Resonators," 39th Annual Frequency Control Symposium, May 1985, pp. 372-374.
3. M. R. Stiglitz, "Dielectric Resonators: Past, Present and Future," Microwave Journal, July 1981, pp. 19-35.
4. "Magnetoelastic Properties of Yttrium-Iron Garnet," Chapter 14, Physical Acoustics, W. P. Mason, Ed., pp. 211-268.
5. E. G. Spencer, et al., "Temperature Dependence of Microwave Acoustic Losses in Yttrium Iron Garnet," Physical Review, Vol. 125, No. 6, March 15, 1962, pp. 1950-1951.
6. S. V. Krishnaswamy, et al., "Oriented ZnO Films for Microwave Shear Mode Transducers," 1982 Ultrasonics Symposium Proceedings (IEEE Cat. No. 82CH1823-4), pp. 476-479.

USING MULTIPLE REFERENCE STATIONS TO SEPARATE THE VARIANCES OF NOISE COMPONENTS IN THE GLOBAL POSITIONING SYSTEM

M. A. WEISS and D. W. ALLAN, National Bureau of Standards, Boulder, CO 80303

Abstract

The separation of variance technique has been applied to measurements of a clock against received signals from Global Positioning System (GPS) satellites to separate out various noise components in the system. In this paper we extend the previous work in several ways. First, we show how measurements can be taken from several different locations to obtain estimates of more components of the GPS system, and to obtain better estimates of the components previously studied. We show how to estimate the variances of the following five components: the GPS system clock, the error in the transmitted correction term between the satellite clock and the GPS system clock, propagation noise in the measurement including ionospheric and tropospheric modelling errors, error in the transmitted ephemeris for the satellite, and the local reference clock. We consider the effects of correlations between elements of the data, and analyze the confidence one may have in our estimates in light of these. Finally, this multi-station separation of variance technique is applied to recent GPS data. We discuss new insights into the GPS system that have been learned using this technique.

Introduction

The Separation of Variance Technique as applied to the Global Positioning System has been reported previously (1) using data taken from one location. This paper reports an expansion of that technique using data taken from multiple reference stations and considering the effect of correlations. Using multiple references allows one to separate propagation noise from ephemeris error variances, as well as providing greater confidence in the estimates of the other noise components: the GPS master clock, the error in the satellite clock correction terms, and the satellite vehicle (SV) clocks. From the study of correlation effects we find better ways to use the data for our estimates as well as understand better some of the limitations in the separation of GPS noise components technique.

We first discuss the method employed for taking data. Next we consider the possible variances one may compute using this data and their components. There are two kinds of variances we discuss: Allan variances of data types, and Allan variances from applying the "N-corner hat" technique to Allan variances of differences of data. Thirdly we discuss the correlation terms contained in these computed variances. Finally we show how to solve these equations for individual noise components in a way which minimizes the effect of correlations as well as indicating the magnitude of some of these correlations. We then apply this technique to recent data, January and February of 1986, and discuss the results.

Contribution of the National Bureau of Standards,
not subject to copyright.

Data

There are two kinds of time measurements we obtain from a GPS receiver: (GPS-Ref)', a measurement of the GPS system clock against the local reference clock with various kinds of noise, and (SV-Ref)', a measurement of the particular GPS satellite vehicle clock against the local reference clock with noise. In practice the (SV-Ref)' is obtained from direct measurements of the received signal against the local reference clock, then corrected by the GPS receiver to account for the motion of the satellite vehicle (SV) using the ephemeris as transmitted from the SV in the real time bit stream. The (GPS-Ref)' value is obtained by decoding the SV clock correction term from the bit stream, and adding it to the (SV-Ref)' value along with a relativistic correction. We have

$$(GPS-Ref)' = GPS + Clx' + Prop + Eph - Ref,$$

where

GPS = the GPS master clock time variation,
Clx' = error in the transmitted correction term between the SV clock and the GPS system clock,
Prop = propagation noise in the measurement including ionospheric and tropospheric modelling errors, receiver noise, and local coordinate errors,
Eph = error in the ephemeris for the SV as transmitted,
Ref = the local reference clock time variation

Also:

$$(SV-Ref)' = SV + Prop + Eph - Ref,$$

where

SV = the satellite vehicle clock phase variation

and where other terms are defined as above.

We obtain these two numbers by tracking and averaging a particular satellite for 13 minutes: (GPS-Ref)' and (SV-Ref)'. Since the satellites are in 12 hour sidereal orbits, we may repeat this measurement one sidereal day later and maintain the same geometry. This is important since the control segment's estimates and uploads of system parameters are all tied to the sidereal period of the orbits. In this way we obtain a time series whose noise components are well defined. The Clx' and Eph terms, as well as the Prop term, to the extent it reflects ionospheric modelling errors, are all defined as errors in the control segment's estimates. We can solve for the variance of these only because our measurements are made with the period of the system being one sidereal day, the same for the control segment and for the users.

Computed Variances

Allan Variances of Data Types

Since we have two time series, the (GPS-Ref)' and (SV-Ref)' data, we may compute three independent Allan variances: the variance of each data type and the variance of their difference. Let us denote these fractional frequency variances as

AG = the variance of the (GPS-Ref)' data type,

AS = the variance of the (SV-Ref)' data type,

AGS = the variance of the (GPS-Ref)' - (SV-Ref)' data,

which is simply the variance of the transmitted clock correction term.

Let us denote the variances of the noise components as

G = the fractional frequency variance of GPS,

C = the fractional frequency variance of Clx',

S = the fractional frequency variance of SV,

P = the fractional frequency variance of Prop,

E = the fractional frequency variance of Eph,

R = the fractional frequency variance of Ref.

With this notation we may write the variance of each data type in terms of the variances of the components. The variance of a sum or difference of random variables is the sum of the variances plus cross-correlations. We will consider correlations later. For now let us assume independence of the components, and we have

$$AG = G + C + P + E + R$$

$$AS = S + P + E + R$$

$$AGS = G + C + S$$

From these three equations we may solve for G+C, S, and P+E+R. If the characteristics of the reference clock are known, we may estimate R independently yielding an estimate of P+E.

The other variances we compute are those comparing differences from at least three SV's. To describe these we first discuss the "N-corner hat" technique.

N-Corner Hat Theory

The separation of variance technique grew out of the desire to know the stability of a particular clock given the fact that measurements of clocks must be made in pairs. The "three corner hat" technique (2) has been used, where pair wise measurements are made among three clocks and the fractional frequency variances of the individual clocks can be found under the assumption of independence of the noise processes of the three clocks. This works as follows:

Let X_{ij} be the time difference measurement clock i minus clock j. Then if the S_{ij}^2 is the variance of the time series X_{ij} , under the assumption of independence, we have

$$S_{ij}^2 = S_i^2 + S_j^2,$$

where S_i^2 is the variance of clock i alone.

In this way we find

$$S_i^2 = (S_{ij}^2 + S_{ik}^2 - S_{jk}^2)/2.$$

Note that, due to finite data sets and correlations in the data, estimates of variances here can be negative. This can usually be taken to imply that the true level of that clock is significantly below the levels of the worst clock.

This technique can be expanded to allow for N-clocks (3). This is an improvement since the above is an exact solution only if there is perfect independence. In practice, because of a finite data length there is some apparent correlation between clocks even if they are physically independent. Also, there are often mechanisms for real correlations. We will discuss later how these appear in GPS data. The "N-corner hat" is defined as a least squares estimate. We minimize:

$$A = \sum_{j=2}^N \sum_{i=1}^{j-1} (S_{ij}^2 - S_i^2 - S_j^2)$$

in order to obtain the solution

$$S_i^2 = (\sum_{j=1}^N S_{ij}^2 - B)/(N-2),$$

where

$$B = (\sum_{i=1}^N \sum_{j=1}^N S_{ij}^2)/(2(N-1)).$$

Note that here, again, estimates of variances can be negative.

N-Corner Hat Computed Variances

We apply the N-corner hat technique to the problem of separating GPS noise components first by differencing our time series between pairs of SV's, and later by differencing our data between pairs of locations. Let us denote our (GPS-Ref)' data via SV "i" minus SV "j" as (GPS-Ref)'_{ij}, and similarly (SV-Ref)'_{ij} and their difference (GPS-SV)'_{ij}, this last being simply the difference of the transmitted clock corrections from the two SV's. In terms of our previous notation, and continuing our use of "ij" for data via SV "i" minus SV "j" we have

$$(\text{GPS-Ref})'_{ij} = \text{Clx}'_{ij} + \text{Prop}_{ij} + \text{Eph}_{ij} ,$$

$$(\text{SV-Ref})'_{ij} = \text{SV}_{ij} + \text{Prop}_{ij} + \text{Eph}_{ij} ,$$

$$(\text{GPS-SV})'_{ij} = \text{Clx}'_{ij} - \text{SV}_{ij} .$$

We see that the clocks in common across different SV's, i.e. the GPS and the reference clock, cancel when we difference the data. This is true exactly only if the measurements are taken simultaneously. If they are not, as is true in the NBS case, then there is a constant time offset in the above phase data from both clocks plus a random fluctuation due to any stochastic nature of the two clocks. These terms can be neglected when we take variances of the data if the noise of the clocks is small enough over the time intervals between tracks. Since the reference clocks we use are those in primary time standards labs around the world this is true at the nanosecond level in our case. Also we find the GPS clock is good enough to allow time intervals between tracks up to 1/3 to 1/2 day.

If we take the Allan variance of each of these time series we obtain variances of the sum of the components plus any cross-correlation terms. Since we will consider correlations later let us assume independence for now. Then we simply have a sum of variances, each being the variance of the difference of a single noise component between two SV's. These we may separate using N-corner hat, again under the assumption of independence, if we have differences among at least three SV's. For example if we start with

$$\text{AGS}_{ij} = \text{the Allan variance of } (\text{GPS-SV})'_{ij} \text{ data,} \\ = C_{ij} + S_{ij}$$

we may use the N-corner hat technique to solve for

$$\text{NGS}_i = C_i + S_i .$$

Similarly, we may solve for NG_i and NS_i by using the N-corner hat technique on Allan variances of $(\text{GPS-Ref})'_{ij}$ and $(\text{SV-Ref})'_{ij}$ data respectively. We may now suppress the index "i" to see what computed variances are available for a specific track of an SV. Thus in addition to the AG, AS, and AGS terms listed above we have

$$\text{NG} = C + P + E ,$$

$$\text{NS} = S + P + E ,$$

$$\text{NGS} = C + S .$$

Here we see we could separately solve for C, S, and P+E. Since we saw before when we considered AG, AS, and AGS that we could solve for G+C, S and P+E+R, by using data from several satellites and the N-corner hat we can separate G from C and P+E from R, and we have redundancy in estimating S. Unfortunately, we shall see later that there are correlation terms to consider which effectively remove the redundancy. Also, we find the confidence of the estimate in N-corner hat depends on the relative size of the noise components. This makes it difficult to estimate R, the reference clocks, if they are significantly quieter than the SV clocks. Before we look in detail at these ideas, we show how we may use common view data from several locations to separate P from E.

We go back to consider (SV-Ref)' data, but now we look at how we can use this data when we have the same SV measured against several locations tracking simultaneously. If we now subtract (SV-Ref)' data taken at location "A" from data taken in common view at location "B" we find

$$(\text{SV-Ref})'_{AB} = \text{Prop}_{AB} + \text{Eph}_{AB} - \text{Ref}_{AB} .$$

We see that the SV clock cancels. Further, the Eph_{AB} term is the differential ephemeris error which tends to cancel dependent on the baseline between locations (4). If we compute the variance of this common view data and apply our N-corner hat technique we find that the original ephemeris variance is reduced by products of common view cancellations (see appendix). The result is that, if we solve for a particular location from N-corner hat, the variance of the ephemeris error is reduced by at least an order of magnitude, which effectively makes it negligible compared to the other terms. If we denote this variance for a given location as NL, we have:

$$\text{NL} = P + R .$$

Since we have estimated P+E+R previously, we may use this to separate E from P+R.

Where Correlations Occur in the Computed Variances

Let us now consider the effect of correlations between various noise components in the system. The physical clocks we believe to be statistically independent. The elements which are estimated by the GPS control segment, however, can have correlations dependent on the ways in which they are estimated. The estimates of SV clock and ephemeris are made based on measurements of signals of the satellites against ground station clocks which are referenced to the GPS master clock. Thus there should be correlations among noise components we have denoted Eph, Clx', and GPS. In practice, the ephemeris is estimated in advance for about 10 days. The GPS real time Kalman Filter then estimates the clock correction value and corrects the long term estimate of ephemeris. For this reason we expect the Clx' and Eph terms to be more highly correlated than other terms, and the Clx' and GPS to be more correlated than Eph and GPS. Let us denote these various cross variance terms as G,C; G,E; and C,E corresponding to the correlations between GPS and Clx', GPS and Eph, and Clx' and Eph, respectively. Then we have

$$\text{AG} = G + C + P + E + R + G,C + G,E + C,E ,$$

$$\text{AS} = S + P + E + R ,$$

$$\text{AGS} = G + C + S + G,C ,$$

$$\text{NG} = C + P + E + C,E ,$$

$$\text{NS} = S + P + E ,$$

$$\text{NGS} = C + S ,$$

$$\text{NL} = P + R .$$

We have 7 equations in 9 unknowns. If we estimate R separately this does not improve, since the equation for NS separates nothing more from AS than precisely R. Thus, if we remove R from the equations above NS

provides no new information and we have 6 equations in 8 unknowns.

Solutions

In practice what we need to do in order to obtain the variances computed above is to select tracks from each location so that each track is made in common view with at least three locations, and each location tracks every satellite at least once. We also estimate the Allan variance of each reference station clock. Then we need to select special tracks so that for each location we have a set composed of exactly one track for each SV. For each location we may then compute NG and NGS using the set of special tracks (and NS if we want to estimate the reference clocks), as well as AG, AS and AGS for each of these tracks. And, finally, for each of these tracks we compute NL using common view data from all the locations involved for the track. We solve for the following Allan variances of components for each of these special tracks, for each location

$$\begin{aligned} G + G,C &= AGS - NGS, \\ C + G,E/2 + C,E/2 &= (AG - AS - AGS)/2 + NGS, \\ S + G,E/2 - C,E/2 &= (-AG + AS + AGS)/2, \\ E + G,E/2 + C,E/2 &= (AG + AS - AGS)/2 - NL, \\ P &= NL - R, \\ G,E &= AG - AGS - NG + NGS. \end{aligned}$$

We see that we solve for the five unknown noise components and one correlation term leaving two correlation terms still affecting our solutions. We note that we could remove the G,E term from our estimates of C, S and E. We choose not to for the following reason. Since the Eph term is a vector error and each track is made from at least three locations, and often these are over large baselines, the Eph term will enter with opposite sign for distant locations. To the extent we have an orthogonal look at the satellite, the G,E and the C,E terms should cancel when we average them over all locations. If we could subtract the G,E term from all of these, we would and attempt to average out only the C,E values. But remember, the above equations can only be used for the special set of one track for each SV for a given location so that we could compute NG and NGS. Thus it is only for these special tracks that we can estimate G,E. However we estimate $G + G,C$ for each of these tracks for each location and average to improve the estimate. As we average tracks with different SV's the G,C correlation is different. It affects our final estimate of G only as an average correlation between the GPS master clock and all clock correction errors.

For the rest of the tracks we compute AG, AS, AGS, and NL. We use our estimated $G + G,C$ also, but only averaged over the special track at each location for a given SV. We solve for the following Allan variances of noise components with possible correlations

$$S + G,E/2 - C,E/2 = (-AG + AS + AGS)/2,$$

$$C + G,E/2 + C,E/2 = (AG - AS + AGS)/2 - (G + G,C),$$

$$E + G,E/2 + C,E/2 = (AG + AS - AGS)/2 - NL,$$

$$P = NL - R.$$

The correlation terms enter here exactly in the form as for the special tracks above where we can use the NG and NGS terms. Thus we may average them, using the orthogonality of our measurement base to cancel some of their effect. We may check this by looking at the $G,E/2 + C,E/2$ residuals after averaging, and then subtract the G,E estimates for those that are among the special tracks.

Thus, we believe we have good estimates of the Allan variances of the GPS system clock, each individual SV clock, the clock correction error for each SV clock, and the ephemeris error for each SV. In addition the Allan variance for these components can be seen as a function of time of the sidereal day, thus allowing one to look for variations in noise components as a function of orbital position. We also have an estimate of propagation noise for each track of each SV from each location. These may be combined in various ways to look for different aspects of the propagation noise. It is a combination of both ionospheric and tropospheric modelling errors and turbulence, as well as multipath effects, coordinate errors at the antenna and receiver noise.

Results

This technique was applied to data taken over the period January 2 - February 27, 1986. Our computer program limited our study to 6 SV's measured from 9 locations. We studied SV's 6, 8, 9, 11, 12, and 13 (Navstars 3, 4, 6, 8, 10, and 9, respectively). The reference stations were the National Bureau of Standards (NBS) in Boulder, Co, the Jet Propulsion Laboratory's Deep Space Tracking Station at Goldstone, Ca (JPL), the National Research Council (NRC) in Ottawa, Canada, the U.S. Naval Observatory (USNO) in Washington D.C., the Paris Observatory (OP) in Paris, France, the Physikalisch-Technische Bundesanstalt (PTB) in Braunschweig, West Germany, the Tokyo Astronomical Observatory (TAO) and the Radio Research Laboratory (RRL), both in Tokyo, Japan, and the NBS radio station WWVH in Kauai, Hawaii. Thus, we have two stations in each of the following areas of the globe: West North America, East North America, Europe, and East Asia, and one station in Hawaii. This provided coverage of all satellites throughout the day with redundancy. The SV's were tracked simultaneously at NBS, JPL, NRC, and USNO, then later at NRC, USNO, OP and PTB, continuing around to OP, PTB, TAO, and RRL then to being tracked by TAO, RRL, and WWVH or TAO, RRL, NBS, and JPL. There were also combinations involving NBS, and JPL with OP and PTB or combining TAO, RRL, WWVH, NRC and USNO. In all there were 122 tracks per sidereal day taken from the 9 sites, all taken in common view among at least 3 sites with a total of 28 different common view track times per sidereal day. There was much information in the output concerning the SV's, the ground stations, and the GPS in general. We discuss some of it here.

First, in figure 1, we see the square root of the Allan variance, the Allan deviation, of the GPS master clock. This is the root mean square (rms) of all estimates taken over the 6 special tracks (1 for each SV) at each of the 9 locations. This behavior at a level of a part in 10^{13} is higher than one would expect from that clock in a good environment.

Figure 2 compares the levels of the three Block I satellites with rubidium clocks. The frequency drift was removed from each of these using a mean second difference estimator. The values reflect rms of estimates over all tracks of each SV at each location. The level we see for SV# 8 is typical of all three when studied over a year, and reflects the lack of constant linear drift over the period in question. This can be seen in figure 3 where we show the phase plot of SV# 8 against NBS with a drift removed. We see that this was a quiet period for SV's 6 and 9. Figures 4, 5, and 6 give more detailed information concerning these three spacecraft. We put the performance of the SV clock along with our estimates of clock correction error and ephemeris error variances on the same plot. Ideally, the clock correction error and ephemeris error levels should be somewhat below the noise level of the clock. This is because it is measurements against the clock that are used to make these estimates, and the redundancy of the measurements should bring the estimates below the clock noise. We see in all three cases that the clock correction error is somewhat above this ideal at one day of integration time. For the ephemeris, however, we see excellent behavior.

Next we look in figure 7 at our estimates of the clocks aboard SV's 11, 12, and 13 (Navstars 8, 10, and 9). SV# 11 was at that time using its 0.1 degree temperature controlled Rubidium clock, while SV's 12 and 13 were Cesium clocks. The performance level we see is consistent with other estimates. We note, however, that SV# 13 has an increase in variance at the two day integration time. This suggests a periodic behavior in the SV phase with a period of about four days. A study of phase plots suggests this is the case, though a cause is unknown. Figures 8, 9, and 10 give the SV clock noise levels on the same plots as the clock correction error and ephemeris error for these SV's, 11, 12, and 13, respectively, as before for the other SV's. Again we see that the clock correction error is somewhat less than ideal, while the ephemeris error level is excellent.

In this run of multi-station separation of variance we also examined cross correlation effects. In solving for the C,F and residual C,E terms we expected to see a reversal of sign for locations widely separated on opposite sides of a satellite. This occurred to some extent only for integration times of 1 day. Data taken at NBS and JPL in West North America and OP and PTB in Europe worked well for analyzing this effect. We conclude that the level of these correlation terms are below the confidence of our estimates for integration times longer than one day. Since we expect the C,E term to be among the highest correlations, this suggests that correlations in the system are not corrupting our estimates. Our limitations for now are the finite data length, and the use of N-corner hat which employs differences of variances.

Other results of note include comments on the estimates of reference clocks, the propagation noise, and on the behavior of the GPS over Asia. We found we were unable to estimate the behavior of the reference clocks, except for the clock at WWVH whose stability was in the 10^{13} level. In general we find that reference clocks can be estimated only to the level of the SV clocks. Our estimates of the propagation noise showed us that some of the locations in our ensemble are slightly noisier than the others. This could be due to multipath or antenna coordinate problems, or, as in WWVH, being closer to the equator where the ionosphere has more effect. Finally, by looking at the clock correction error variance as it behaved throughout the day we saw some tendency for it to be worse as satellites were over Asia or the Pacific.

Conclusion

In conclusion we see that this separation of variance technique has grown from being powerful in its inception to a technique providing a wealth of information about many important aspects of the GPS. We have found that the use of multiple references gives us greater confidence in the estimates of the physical clocks in the system, the GPS master clock and the SV clocks, as well as providing better estimates of the GPS Kalman estimation error: the error in the satellite clock correction terms and the ephemeris error variances which we can now in large part separate from the propagation noise. From the study of correlation effects we have found better ways to use the data for our estimates as well as to better understand limitations of the separation of GPS noise components technique.

ACKNOWLEDGEMENT

We wish to thank the Air Force Space Division for partial support in the development of this technique. We also wish to thank Dr. Henry Beisner and Dr. Judah Levine for assisting with the paper's review process.

Appendix: Common View Cancellation of Ephemeris Error

Let us consider cancellation of ephemeris error among 3 sites in common view, the general case among $N > 3$ sites only having better cancellation. Let us label the sites A, B, and C. We already understand that (SV-Ref)' data contains an ephemeris error term which we denoted Eph. But to understand cancellation we must view the Eph term for location A as a vector ephemeris error, \underline{Eph} , projected in the direction \underline{e}_A from the SV to the ground station A. Then

$$\begin{aligned} (SV-Ref)'_A &= SV + Prop_A + \underline{Eph} * \underline{e}_A - Ref_A \\ \text{and} \\ (SV-Ref)'_B &= SV + Prop_B + \underline{Eph} * \underline{e}_B - Ref_B, \end{aligned}$$

so

$$(SV-Ref)'_{AB} = Prop_{AB} + \underline{Eph} * (\underline{e}_A - \underline{e}_B) - Ref_{AB}.$$

If we take the Allan variance of this expression we have

$$AS_{AB} = P_A + P_B + (\underline{e}_A - \underline{e}_B) * \underline{E} * (\underline{e}_A - \underline{e}_B) + R_A + R_B,$$

where E is the covariance matrix of the ephemeris error. A simple computation using the linearity of the covariance matrix E shows that we have, in the three-corner hat solution

$$(AS_{AB} + AS_{AC} - AS_{BC})/2 = P_A + (e_A - e_B)^* E^* (e_A - e_C) + R_A.$$

The expression

$$(e_A - e_B)^* E^* (e_A - e_C) \leq (e_A - e_B)^* (e_A - e_C)^* E_{\max},$$

where E_{\max} is the maximum ephemeris error in an orthogonal coordinate system. This follows since, with E symmetric, we can choose a coordinate system which diagonalizes it. Thus the ephemeris error is reduced in the variance not simply by the common mode cancellation term $(e_A - e_B)$, but by the dot product of a pair of cancellation terms! In the general case for N locations this expression becomes an average of pairs of cancellation terms. The worst possible case for this term occurs when 3 ground stations are located 120 degrees apart around a great circle with the SV on the orthogonal axis. In that case we find

$$(e_A - e_B)^* (e_A - e_C) < 0.2.$$

Of course, this case cannot occur in practice since the SV would be below the horizon at all sites. Thus we see that the variance of the ephemeris error cancels to at least an order of magnitude in the N -corner hat computed variance across locations, NL .

References

1. "Separating Some of the Error Sources in the Global Positioning System by Variance and Linear Analysis," David W. Allan and Marc A. Weiss. Accepted for publication in Radio Science.
2. "A Method for Estimating the Frequency Stability of an Individual Oscillator," James E. Gray and David W. Allan Proc. 28th Annual Symposium on Frequency Control, Fort Monmouth, NJ, (1974), p.243. (Available from National Technical Information Service, Sills Building, 5285 Port Royal Rd., Springfield, VA 22161).
3. "Time Scale Algorithms Using Kalman Filters - Insights from Simulation," James A. Barnes. 2nd Symposium on Time Scale Algorithms, 23-25 June, 1982, NBS, Boulder, appendix C.
4. "Accurate Time and Frequency During Common View of a GPS Satellite," David W. Allan and Marc A. Weiss. Proc. 34th Annual Symposium on Frequency Control, Fort Monmouth, NJ, (1980), p.334. (Available from Annual Frequency Control Symposium, c/o Electronic Industries Assoc., 2001 Eye St., Washington, DC 20006).

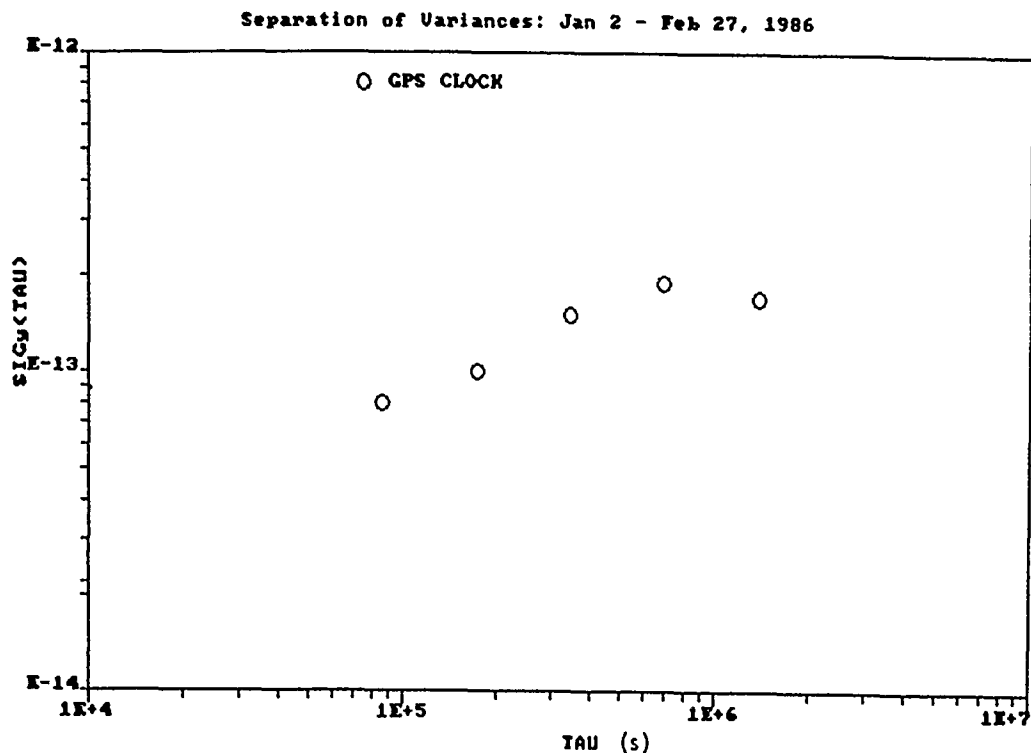


Figure 1: We see the square root of the Allan variance, the Allan deviation, of the GPS master clock. This is the root mean square (rms) of all estimates taken over the 6 special tracks (1 for each SV) at each of the 9 locations. This behavior at a level of a part in 10^{13} is higher than one would expect from that clock in a good environment.

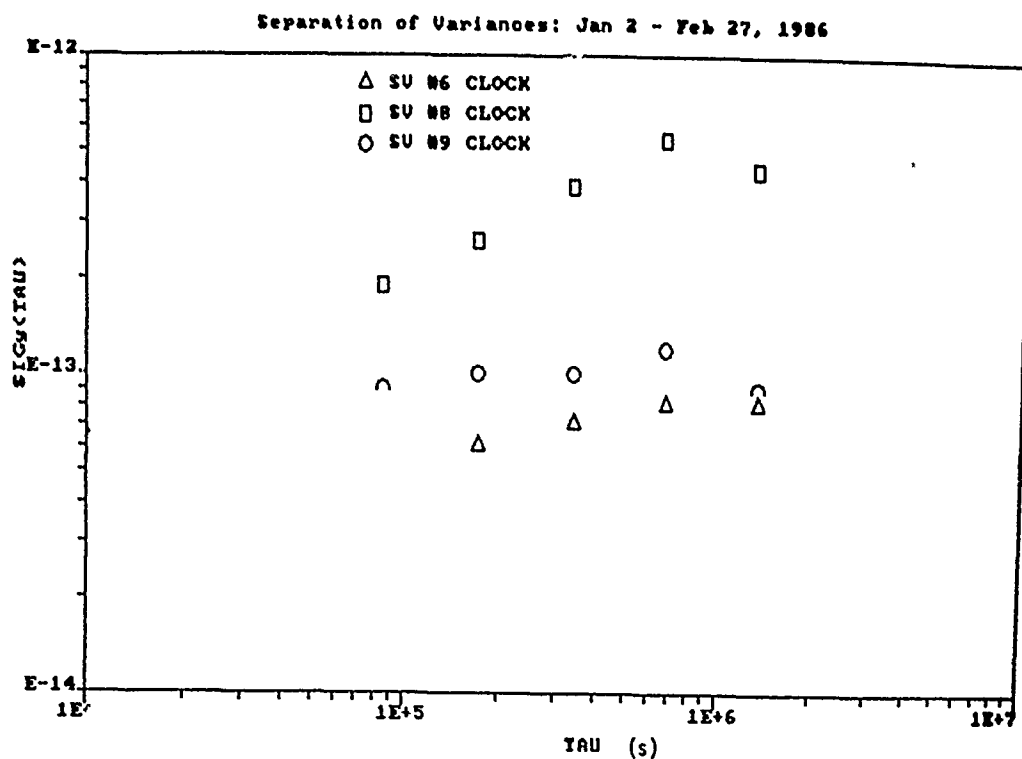


Figure 2: The levels of the three Block I satellites with rubidium clocks. The frequency drift was removed from each of these using a mean second difference estimator. The values reflect rms of estimates over all tracks of each SV at each location. The level we see for SV# 8 is typical of all three when studied over a year, and reflects the lack of constant linear drift over the period in question.

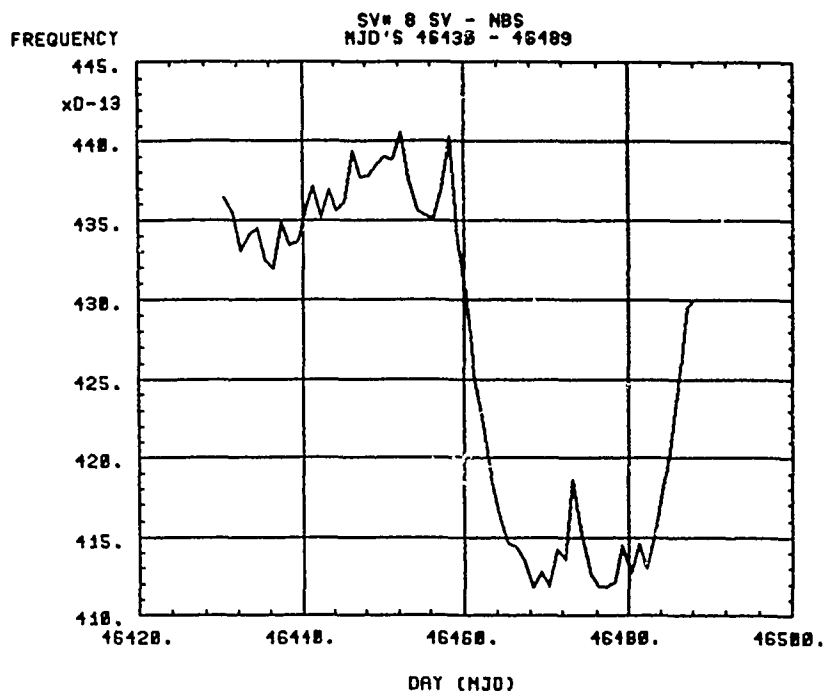


Figure 3: The phase plot of SV# 8 against NBS with a drift removed showing the lack of a constant drift over the period of this analysis.

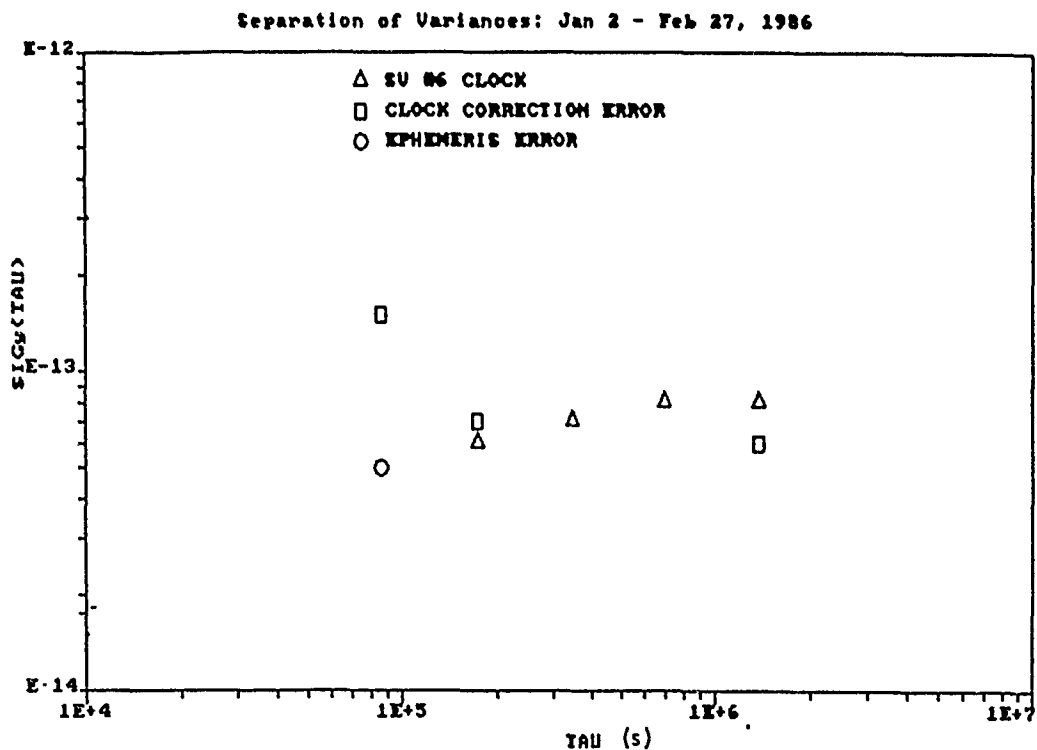


Figure 4: The noise or error levels of SV# 6 (Navstar 3) components. The clock correction error level is worse than the noise level of the SV Rubidium clock at one day.

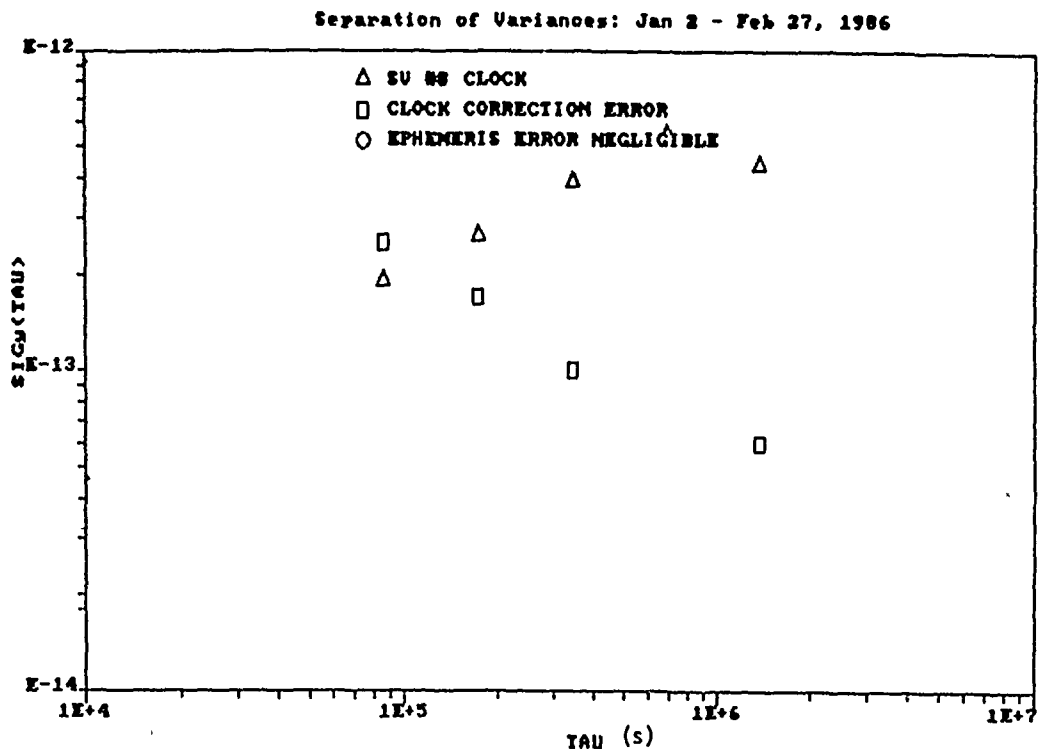


Figure 5: The noise or error levels of SV# 8 (Navstar 4) components. The clock correction error level is worse than the noise level of the SV Rubidium clock at one day.

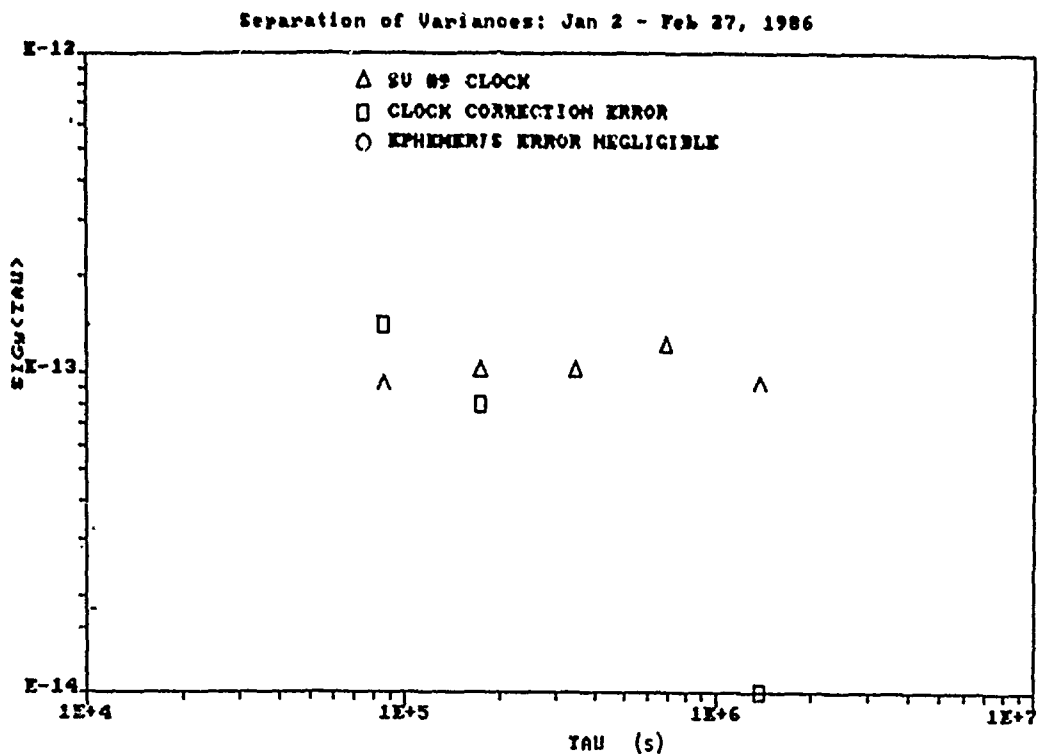


Figure 6: The noise or error levels of SV# 9 (Navstar 6) components. The clock correction error level is worse than the noise level of the SV Rubidium clock at one day.

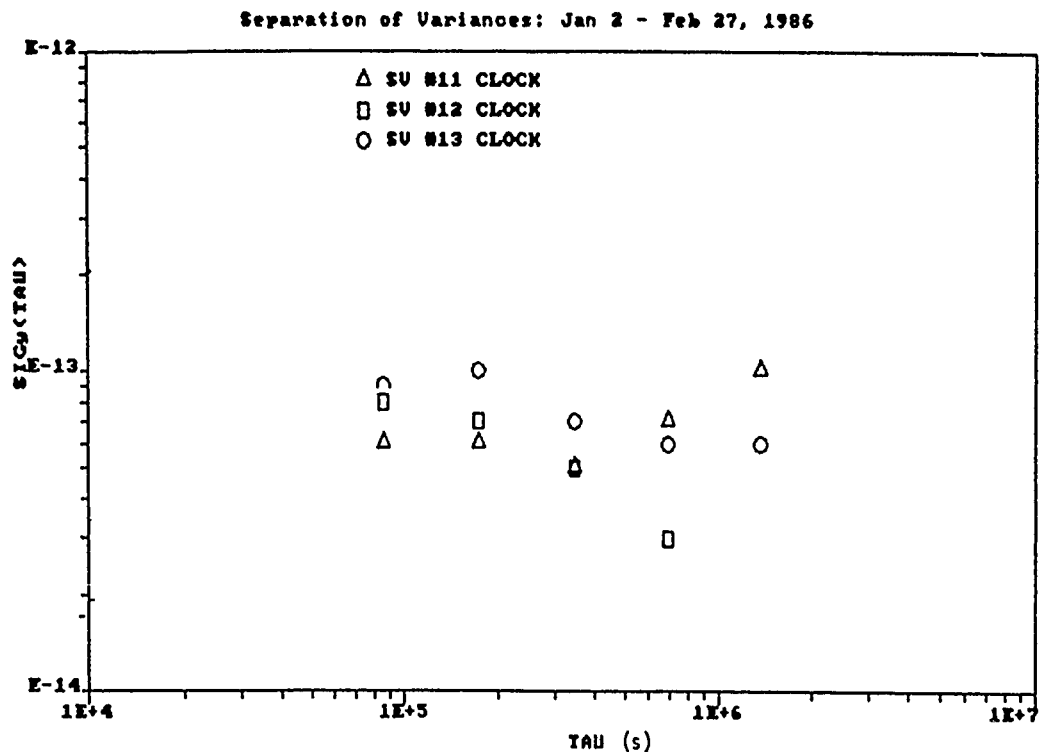


Figure 7: Our estimates of the clocks aboard SV's 11, 12, and 13 (Navstars 8, 10, and 9). SV/ 11 was at that time using its temperature controlled Rubidium clock, while SV's 12 and 13 were Cesium clocks. The performance level we see is consistent with other estimates.

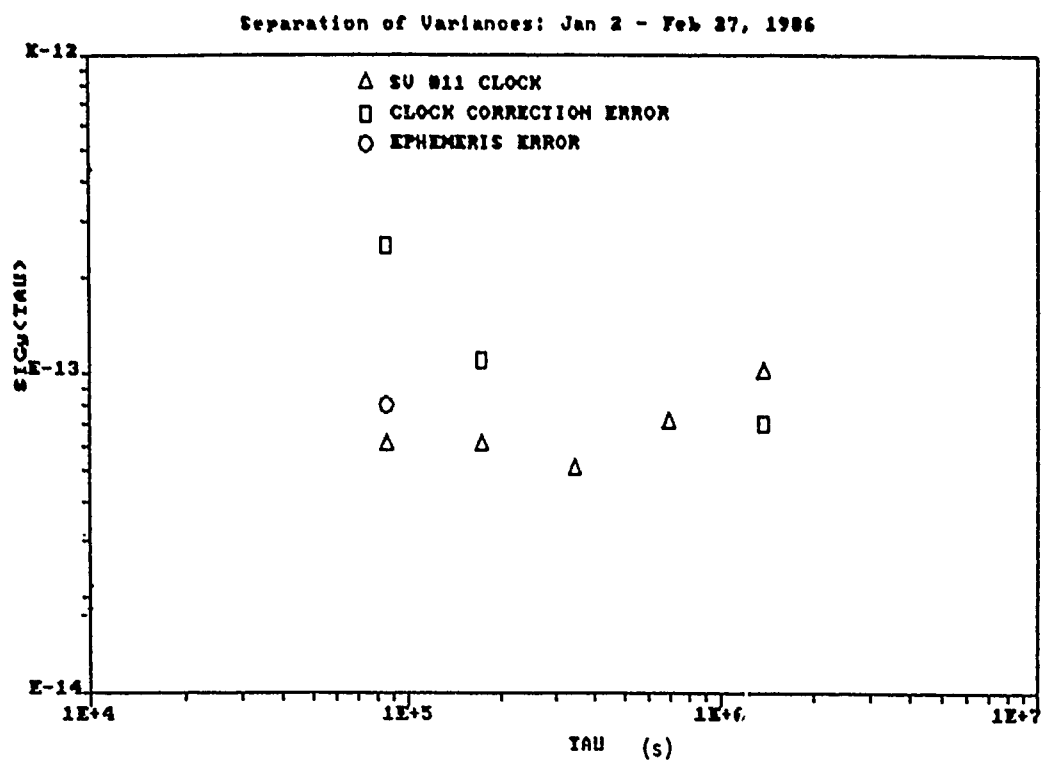


Figure 8: The noise or error levels of SV# 11 (Navstar 8) components. The clock correction error level is worse than the noise level of the SV temperature controlled Rubidium clock at one day.

Separation of Variances: Jan 2 - Feb 27, 1986

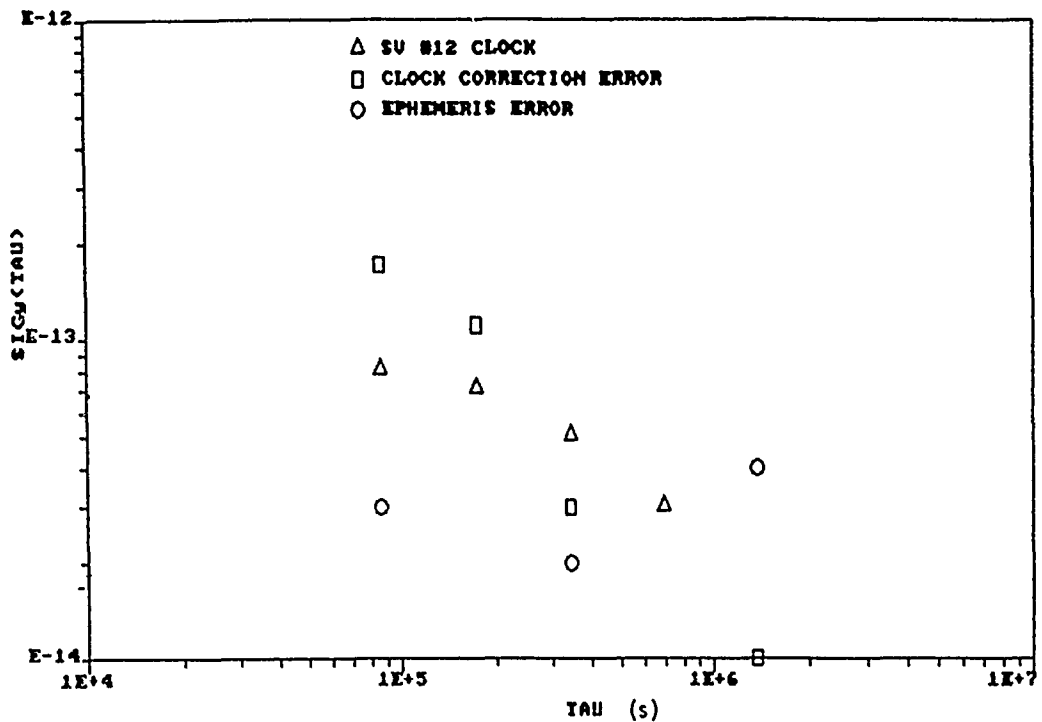


Figure 9: The noise or error levels of SV# 12 (Navstar 10) components. The clock correction error level is worse than the noise level of the SV Cesium clock at one day.

Separation of Variances: Jan 2 - Feb 27, 1986

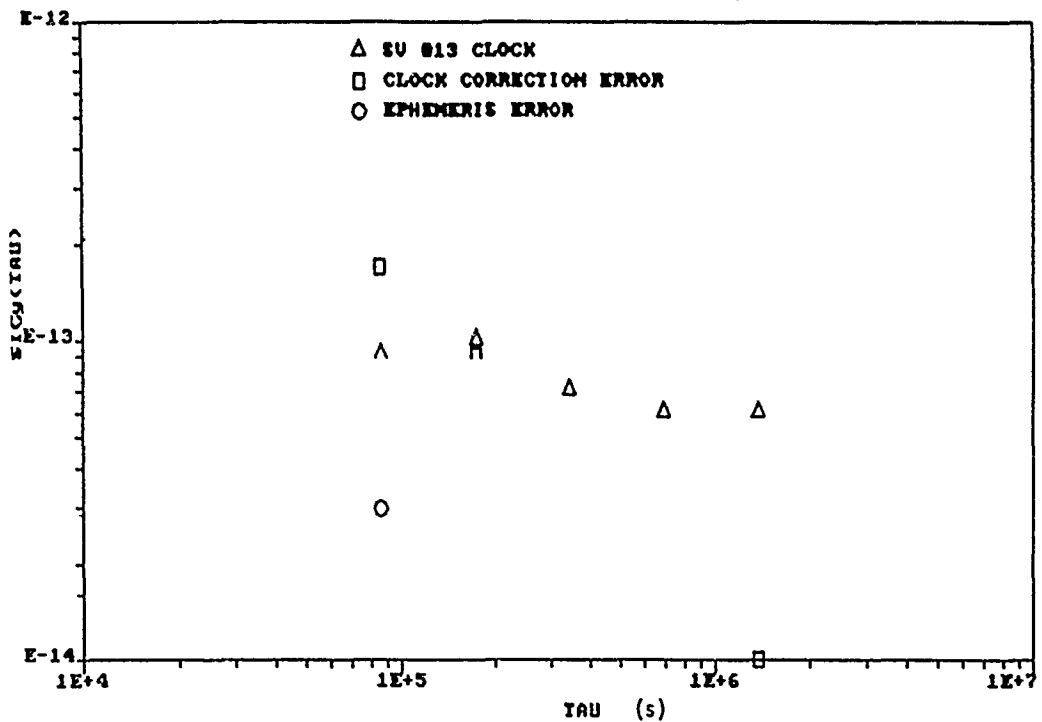


Figure 10: The noise or error levels of SV# 13 (Navstar 9) components. The clock correction error level is worse than the noise level of the SV Cesium clock at one day.

THE U.S. NAVAL OBSERVATORY (USNO) PTTI DATA SERVICE

by

Gernot M. R. Winkler
U.S. Naval Observatory
Washington, DC 20390-5100

ABSTRACT

For the last 5 years direct access to one of the USNO's computers has allowed users to obtain precision time data in near real time. This includes status information for the major timed systems such as downtimes and announcements of planned frequency adjustments. USNO press releases, Stargazer's Bulletins and general information files concerning Precise Time and Time Interval (PTTI) systems are also available. The more important files are updated daily on the basis of USNO's extensive remote data collection, and the message traffic received. The service is available "around the clock" and conforms with U.S. modem standards as well as CCITT at 300, 1200, and 2400 Baud (even parity). The experience gained in the operation of this service has suggested a substantial further development in the data collection and data dissemination efforts in support of PTTI operations. These improvements have established a system for extensive real time PTTI information for users anywhere. This system is described and plans for a future extension are sketched. This extension is to eventually produce part of a "Repository of PTTI information" as directed in the latest DoD PTTI instruction. The system will also replace most of the current bulletins now distributed by mail.

INTRODUCTION AND BACKGROUND

The USNO has provided a scientific data service since the middle of the nineteenth century. The more important publications of the USNO are well known to the nautical and astronomical communities. We publish annually the Nautical Almanac, the Astronomical Almanac (formerly the American Ephemeris), the Air Almanac, the Astronomical Phenomena, a variety of Circulars such as the solar eclipse circulars, in addition to supporting information. In 1977, the Nautical Almanac Office also introduced the Almanac for Computers which meets the requirements of those who use small desktop computers or calculators to obtain their data. All these publications are available either from the U.S. Government Printing Office (such as the Almanacs) or directly upon request from the USNO (such as the Circulars).

After the observatory began the dissemination of precise time via radio (1904), the increased use of this radio time brought about the need for more detailed information on frequency and time adjustments of these radio time signals in the form of Time Service Bulletins. In the 1950's the precise frequency control of Naval Very Low Frequency (VLF) radio stations provided a more accurate means of time dissemination than the conventional HF time signals. In the 1960's, the timing of LORAN C was the next step, quickly followed by the use of the timed navigation system TRANSIT. All of these operations also increased the need for more information to the general user community and as a result, the Time Service Bulletins changed and grew. Perhaps an even more pressing need for up-to-date scientific data services originated with the astronomy-geodetic-space sciences community which needs accurate data on Earth orientation (Polar Motion, Universal Time) for satellite tracking, pointing, and orbit prediction.

By the late 70s more rapid means for the dissemination of service bulletins became available in the form of direct telephone access to one of the computers in the observatory. It was soon apparent that those users who had the most demanding and critical applications preferred this method of data transfer over the much slower mail distribu-

tion. Commensurate with the extension of our internal and external data collection, the handling and storage of our reference data also underwent a complete revolution. Instead of first producing the bulletins on paper and then entering them into the computer, almost everything is now done the other way around. For this reason, the files which are actually published in bulletin form constitute only a small subset of the data which are directly accessible by phone. Moreover, the cost of the production and mailing of bulletins is becoming a real burden which we want to reduce by relying more and more on electronic means of data dissemination. We actually have firm plans to cease production and mailing of the majority of our bulletins in the foreseeable future. This report, therefore, has as its aim the announcement of this policy, the encouragement for wider utilization of the electronic data service, and the solicitation of suggestions for further improvements in scope and availability.

THE USNO HIERARCHICAL DATA COLLECTION AND PROCESSING SYSTEM

Data acquisition, process control, data reduction and storage at the USNO is done in a hierarchical, widely-dispersed computer system. Figure 1 shows the overall principle of this distributed system. I should add here that the Observatory's operations take place not only at Washington, DC, but also in southern Florida (the alternate time service station), Flagstaff, AZ (Astrometry and observations with CCD's), and the Black Birch Observatory in New Zealand (for Astrometry). In addition, we interface with several dozen Precise Time Reference Stations (PTRS) from which data are obtained regularly. We also receive large amounts of data from various radio astronomy sites.

At the local level (level 1) a microcomputer takes care of the data collection and process control. It also provides buffer storage and the communications interface (hardwire, fiber-optics, or the telephone) to the next higher level. We use an HP 9915 at most places. We are now including PC's with appropriate interfaces as system controllers at the local level. In addition, PC's serve also other purposes. Eventually, everyone in the Time Service Department is going to use a PC as the major tool to obtain, evaluate, and disseminate timing data. Each one of these workstations can access directly one of the major computing resources or sources of data, and can, in turn, make information available to the system or to the wordprocessors for hardcopy production.

Level 2 is occupied by complexes of minicomputers such as the IBM SERIES 1 and the HP 1000 systems. These machines in Time Service itself are dual processor systems for backup capability and decoupling of critical operations. At this level remote and local data are collected from the level 1 processors, instruments are activated and measurements are made. Preliminary data filtering and data reduction is also performed at this level. Figure 2 gives an overall sketch of the role of the level 2 nodes.

Finally, the mainframe system is depicted in Figure 3. Here the emphasis is on very large scale data handling with increasing importance given to image processing (in support of optical and radio observations). The system is about to be upgraded to make more data channels available mainly because of this growing use of image processing. The image processing itself will then take place entirely in dedicated subsystems of the VAX type. This development is driven by

the needs of the radio astrometric observations done with the VLA and Green Bank Interferometer and, more recently, is also necessary in support of the new VLBI correlator (MARK 3) facility at the USNO. The increasing use of CCD arrays directly in the focal plane and also in plate measuring machines produces similar requirements.

The links between these various levels are almost all serial data links. Computer to computer and mainframe remote access at the observatory is handled predominantly by fiber optics. The superior data handling capability of these links is not the only advantage for us. The electric isolation is most important for an installation which is spread out over many buildings. Eventually most of our links will be fiber optic as they are insensitive to lightning strikes and ground currents. For long distance signalling, 1200 Baud modems are standard with an ongoing transition to 2400 Baud. The last installed PTRS at Colorado Springs is equipped with a 2400 Baud modem with error checking and accelerator. Depending on line condition, effective speeds up to 9600 Baud are in use. The benefits of going to higher speed include savings in connect time, but much more important is the fact that a shorter connect time drastically reduces the chances of interrupts and loss of connection during a data transfer. Critical data transfers are handled with redundancy checks for additional insurance of data accuracy.

A substantial part of our data still arrives via the TWX II. In addition, we receive data and make them available via several other services as shown in Figure 3. The G.E. Mark III utilization has become the standard for international data storage and exchange in support of timing and Earth rotation parameter observations.

THE DATA ACCESS SYSTEM (ADS) OF THE USNO AS IT EXISTS TODAY

The telephone access under the ADS is presently handled by the HP1000 F processor. In addition, we offer a few other services such as summarized in Table 1.

TABLE 1

Telephone Time Services:

AUTOMON ACCESS CODE 294-	COMMERCIAL AC 202 653-	
-----	FTS 653-	
	EXTENSION:	
DATA SERVICE (ADS) EVEN PARITY, 1 START, 1 STOP:		
-----	BELL 103/212 (300/1200 BAUD)	1079
-----	CIITT V.21 300 BAUD	1095
-----	CIITT V.22/V.22bis (1200/2400)	1783
EASYLINK MAILBOX 62920428		

TELETYPE: 710 822-1970 (MESSAGES RECEIVED)		
FIVE OF DAY (VOICE)	202-653-1800	
-----	FOR AUTOMON ONLY 294-1920	
	LONG DISTANCE (TOLL) 900-410-TIME	
	(50c FIRST MINUTE)	
TIME OF DAY FOR DIGITAL CLOCKS (1200BD)	202-653-0351	
-----	EVEN PARITY, 7 BIT ASCII FORMAT	
	MJD-DOY-HHMMSS-UTC CRLF*QULF	
	(SEE EXPLANATION FILE QTSF IN THE DAS)	
ISA SERIES 5 (VOICE)	-1757	

The ADS operates with full duplex. However, the echo can be turned off. This is recommended for data uploads. We still use the Vadic modem standard in addition to Bell 212. However, in the course of further developments in our communications system we may in the future want to drop the VADIC 1200 style (VADIC 3400). We have no evidence

that anyone is still using this modem standard. The 1200 Baud which we want to use for U.S. traffic is the BELL 212 standard. The majority of all calls still arrive at 300 Baud (Bell). Very few callers seem to be interested in the 2400 Baud capability even though this standard is a true international standard and allows considerable savings in connect time. This is particularly true for callers who have automated their calls which is very convenient by means of the better communications programs available for the MS-DOS, and CP/M machines (and also for the Apple McIntosh and the Commodore Amiga). A call which utilizes a pre-programmed command sequence ("Script-files") can access the ADS, obtain a couple of status and data files and terminate within a single minute of connect time. A print-out can then be performed from file after the call. The use of completely automated calls offers particular advantages for the busy executive who has trouble hitting the right key. That can become expensive during a toll call!

The protocol is standard ASCII, i.e., Even Parity, 7 data bits, one start, one parity, and one stop bit. The ADS comes on line automatically and requests identification. This ends with an Enquire character prompt. All other "turn around" prompts are Bell, Asterisk, Carriage Return and Linefeed. Provisions exist to turn off the 3 idle characters transmitted after each record on 300 Baud (or to turn them on at 1200). The ADS will interpret all lines as incoming message for the log-file except those lines which start with an @ character (commercial "at", ASCII code decimal 64 or octal 100). This is, therefore, the command character. For help, one uses a question mark. The first level menu is available with @TCO (for table of codes). This file is reproduced here as Table 2.

TABLE 2

ADS FIRST LEVEL MENU

TABLE OF CODES @TCO FOR THE DIGITAL DATA ACCESS SYSTEM:

FOR ANY OF THE FOLLOWING SERVICES USE THE COMMAND AS GIVEN BUT LEFT JUSTIFIED, I.E., THE COMMAND SYMBOL @ MUST BE THE FIRST CHARACTER IN THE LINE. THIS IS THE FIRST LEVEL MENU: MORE CODES CAN BE FOUND UNDER THE RESPECTIVE EXPLANATIONS.

EXPLANATIONS, GENERAL. . . @EXP	
SPECIAL DAILY MESSAGE. . . @LNE	TIME SERVICE DIRECTORY @DIR
GENERAL PTTI MESSAGE. . . @MES	PTTI CONFERENCE NEWS. @TTI
OTHER SOURCES FOR INFORMATION ON THE OPERATIONAL STATUS OF ELECTRONIC NAVIGATION SYSTEMS ARE LISTED IN FILE . . . @NAU	
ANNOUNCEMENT SERIES 4,5 & 7 CODES, INFO & EXPLANATIONS . . . @SEIXP	
NBS INFO CODES & EXPL. . . @NBSXP	
OMEGA CODES & EXPL. . . @ONSXP	
VLF CODES & EXPL. . . @VLEXP	
GPS CODES & EXPL. . . @GPSXP	
TRANSIT CODES & EXPL. . . @TRAXP	
LORAN CODES & EXPL. . . @LORXP	
TV CODES & EXPL. . . @TVIXP	
PORTABLE CLOCK INFO. . . @MFCXP	
TIME OF COINC. (LORAN,TV) @TCC	
REAL TIME MEAS. EXPL. . . @RTMXP	
STANDARD TIMES FOR ALL COUNTRIES, EXPLANATIONS @STTXP	LIST OF USNO PRESS RELEASES, GENERAL INFO. . . @STAXP
TIME TO +-50MS @TIM	TIME SIGNAL EXPL. . . @TSF
EXPLANATION FOR MJD . . . @MILXP	
CONVERSION OF MJD TO DOY @MJD	CONVERSION: DOY TO MJD @DOY
SIDEREAL TIME @STI	DATE, MJD, WEEKDAY . . . @DAT
SUNRISE, SUNSET, TWILIGHT FOR ANY POINT. @SRI	PHOGRAM EXPLANATIONS. . . @SRIXP
OPERATIONS CONTROL: FOR EXPLANATIONS, DETAILS AND CODES SEE @OPSXP	
CONNECTION SHOULD BE TERMINATED WITH CONTROL D, @END OR @BYE	

At this time no special access code is required; the caller is simply asked to give some meaningful identification. This is necessary for the computer because with a fixed identification for each caller, a message can be given in answer to questions or requests.

The first level menu (@TCO) will guide the caller to the more specific menus which exist for every file group. An example would be the GPS files which are listed in file @GPSXP (Table 3).

TABLE 3

THE GPS FILE GROUP

GPS STATUS INFORMATION.	@GPS
THIS FILE IS UPDATED EVERY WORKDAY.	
INFORMATION OF TEMPORARY INTEREST IS IN FILE. . . .	@GPSTU
INFORMATION OF PERMANENT INTEREST IS IN FILE. . . .	@GPSPO
GPS SYSTEM DESCRIPTION.	@GPSSY
Sources for GPS Receivers	@GPSOU
TRACKING DATA ST1 (502) RECEIVER AT NAVOBSY	@GPSD1
TRACKING SCHEDULE FOR THE ST1 (502) RECEIVER. . . .	@GPSD2
(COMMON VIEW SCHEDULE)	
ACTUAL TRACKING OF ALL DATA	@GPSCH
TRACKING DATA FOR THE LAST 7 DAYS	@GPSD3
EXTRAPOLATION COEFFICIENTS FOR MC - GPS CLOCKS. . . .	@GPSD4
(USES THE LAST 5 DAYS)	
EXTRAPOLATION USING LAST 3 DAYS	@GPSD5
SAME AS FILE @GPSD1 BUT IN COORDINATED FORMAT	@GPSD6
SAME AS FILE @GPSD3 BUT IN COORDINATED FORMAT	@GPSD7
SATELLITE HEALTH CODES	@GPSD8
PREDICTION FOR GPS MESSAGE PARAMETERS A0 AND A1	@GPSD9
(GPS FORMAT, 3 DAY AVERAGES TAKEN)	
TRACKING DATA FILE INFORMATION BACK TO Jan 86	@GPSSL
FILE IS VERY LONG. FOR LAST 7 DAYS USE @GPSD3	
SUMMARY OF PAST 15 PREDICTIONS (3 DAY SMOOTHING). . . .	@GPSV1
SUMMARY OF PAST 15 PREDICTIONS (1 DAY SMOOTHING). . . .	@GPST4
EXPLANATIONS FOR THE PREDICTION COEFFICIENTS. . . .	@GPSV2
GENERAL EXPLANATIONS FOR PREDICTIONS.	@GPSV4
FIT COEFFICIENTS FOR LINEAR AND QUADRATIC FITS. . . .	@GPSV5
DATA REDUCTION TO UTC(USNO) WITH NAV.MESS.DATA. . . .	@GPSB6
(COMPARISON COMPUTED, USING A0 AND A1 CONSTANTS, AS READ IN NAV MESSAGE). THE SAME INFO IN DIFFERENT FORMAT IS ALSO IN	@GPSMC
DATA REDUCTION TO UTC(USNO) WITH 7 DAYS OF DATA	@GPSU7
(SAME FORMAT AS FILE @GPSB6).	
FOR DATA FROM TTU PROTOTYPE RECEIVER SEE.	@GPXXP

As one can see, there is a considerable amount of information available on line. All of the files of real time interest are being updated on every work day. A typical example would be the status files for systems such as GPS (@GPS), TRANSIT (@TRA), or LORAN (@LOR). Access to these status files will answer many questions and resolve some problems for users of these systems. However, the ADS can be helpful also in a variety of other cases. Programs are on line for computing distances, LORAN signal delays, TRANSIT visibility, Times of Coincidence (TOC), Sunrise, Earth orientation data predictions (explained in @SERXP), Sidereal time, conversion from date to MJD and reverse, etc. Lastly, you can also find general astronomical news and other pertinent information files in the system. In one word, everything of relevance to PTTI should be in this system, except, naturally, games!

PLANS FOR A PTTI DATA "REPOSITORY" AS DIRECTED IN DOD DIRECTIVE 5160.1

At this time, another major step is being taken: The enlargement of the present system to become part of a "Repository of PTTI information" as directed in the recent DoD Directive 5161.1 (paragraph E.e). The purpose of this "Repository" is clear: There exists a need to assist system planners, designers, operators, and logistic managers in obtaining information on existing assets and available supports in near real time. This data bank should include a guide to more detailed background information concerning all questions of frequency control and timing. We have completed the first step in the project and implementation has started. The "Repository" is going to exist in two parts: A completely open part, available as addition to the above described ADS, and a somewhat restricted part, accessible only with pre-assigned access codes and residing on a separate dedicated machine. This division appears logical because it will allow a combination of two opposing goals, the widest accessibility and the most complete and useful information for official users.

In the open part of the system we envision having lists of sources of equipment (such as the rudimentary beginning in the form of file @GPSOU in table 3), capabilities and explanations of systems, standards, available equipment, training requirements, maintenance support, specifications, literature guides, if not a complete PTTI bibliography, and up-to-date information that is more extensive than is provided at this time. The restricted access part will contain lists of stations and their capabilities, traceability, R&D programs, user requirements, etc.

The system allows a wide latitude of choices and we would very much appreciate any suggestions, ideas, and requirements you may have. It is our intent to make the "Repository" as useful as possible and this can only be accomplished with wide participation of the users themselves. Regarding availability of equipment and PTTI related items we would certainly welcome material from suppliers in a form suitable for inclusion. By far the easiest way for us would be a direct uploading of your material into a dedicated mailbox in the ADS, alternatively, the receipt of a data diskette.

Acknowledgement

I must acknowledge here the essential contributions which have been made to this project by all members of Time Service Department. Not only does the original implementation of such an information system and its development into a well working user aid require a real team spirit, but almost more demanding is the daily effort needed to assure that all pertinent information actually finds its way into the system, reliably, promptly, and accurately. This demands dedication, perseverance, and judgment.

References

1. Miranian, Mihran (1982), "The U.S. Naval Observatory Automated Data Service." Presented at the HP1000 International Users Group Conference 1982, Long Beach, California.
2. Wheeler, Paul J. (1983), "Automation of precise Time Reference Stations", Proceedings 15th Precise Time and Time Interval Applications and Planning Meeting, p. 41-52.
3. Withington, F. N. (1984), "G.E. Mark III RC28 Catalog Users Guide", available on request from USNO.

- 1 LOCAL LEVEL:
 HP 9915 (Serial & IEEE 488 IF), modem 1200/2400 baud
 PC, modem

 - 2 PROCESS CONTROL AND DATA ACQUISITION NODES
 HP 1000 Systems, IBM Series 1, Vax for image processing in Flagstaff

 - 3 Mainframe computer (IBM 4341)
- Links:
- a) Serial data links via short-haul modems (2400 - 9600 baud)

 - b) Serial links via dial-up modems (1200, 2400 baud)

 - c) Fiber optic links to mainframe

THE USNO THREE LEVEL DATA SYSTEM

FIGURE 1

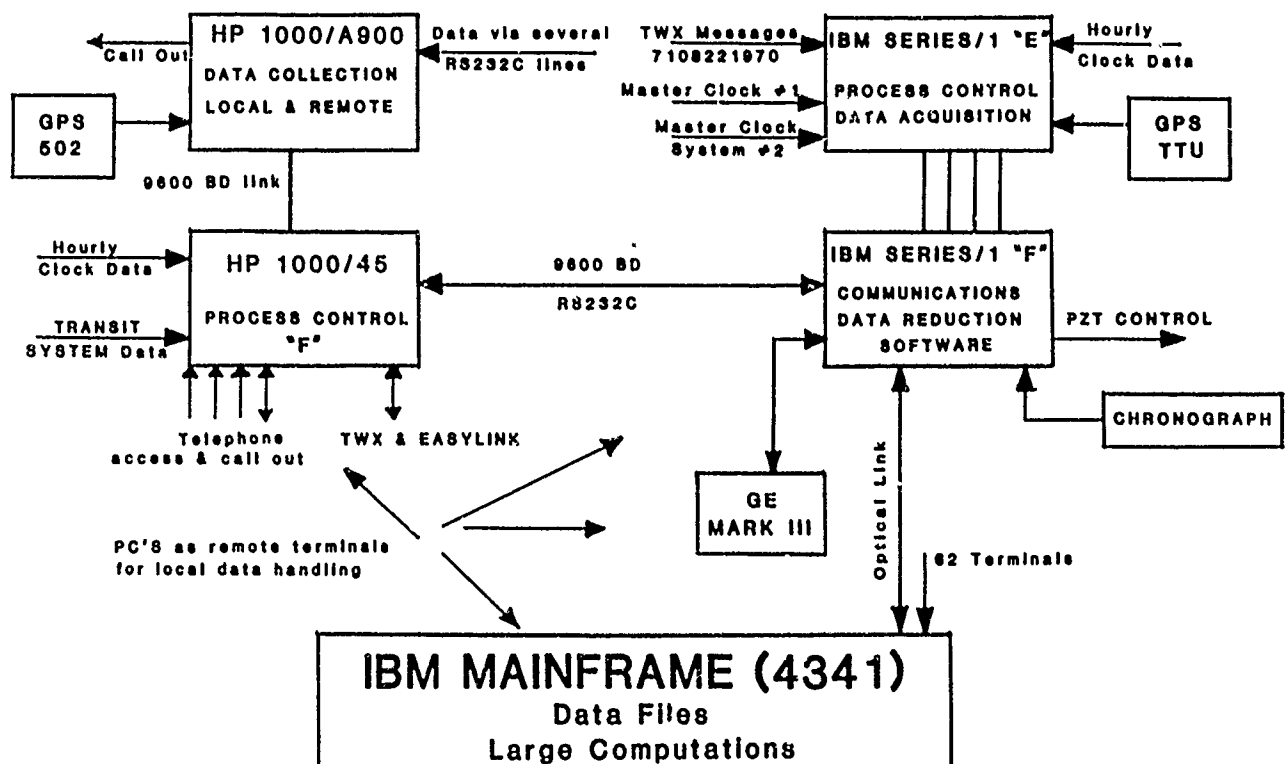


FIGURE 2

IBM 4341 (4381 type)
4 Mbyte

	Serial Links to	Level 2 Nodes
CHANNEL:	"	Plotter, WP, Telephone Access
	"	62 remote terminals and printers
CHANNEL:	Disk and tape storage (12 x 3350 + 4) (8 x 2314) (8 tapes)	
CHANNEL:	Image Processor Data Interface (micro Vax Vax)	

MAINFRAME COMPLEX

FIGURE 3

G. E. MARK III INFORMATION SYSTEM

Catalog RC28 for worldwide data exchange
and filing (see Reference 3)

Science Net (OMNET INC. Tel 617/2659230; using GTE

TELENET)
for VLBI data exchange and coordination
Bulletin Board: VLBI
Address ID USNO.VLBI

Easylink (Western Union)

USNO Mailbox 62 920 428

TWX II (Western Union) 710 8221970

EXTERNAL DATA SERVICES

FIGURE 4

ACTIVE H-MASERS FOR V.L.B.I. APPLICATIONS

R. Barillet, P. Petit, J. Viennet, C. Audoin
Laboratoire de l'Horloge Atomique
Equipe de Recherche du CNRS,
associée à l'Université Paris-Sud
Bât. 221 - Université Paris-Sud
91405 ORSAY - France

Summary

Three active H-masers have been manufactured at L.H.A. for french radioastronomers and geophysicists. They are transportable and compact. Including electronics, power supplies and batteries, their volume is about 0.5 m^3 and their mass is 280 kg. They are operated from 220 Vac or 24 Vdc power supplies, and their internal battery makes it possible to operate them during transportation. Signal receiver and phaselocked loop have been designed to optimize the frequency stability transfer. Operating conditions are microprocessor controlled and monitored. Measurements of the short term frequency stability, and of systematic effects are presented.

The package design is an evolution of the TE_{011} mode cavity masers made at L.H.A.¹, but the new masers are much more compact and light weight.

The use of a single Vac-Ion pump, the replacement of indium alloy O-rings by much softer O-rings, the use of a small thickness (5 mm) quartz cavity cylinder and other tradeoffs enabled the reduction of the overall size of the frequency standard ($H = 117 \text{ cm}$, $L = 77 \text{ cm}$, $\lambda = 53 \text{ cm}$).

The ensemble composed of the microwave cavity and the storage bulb is temperature compensated. The achieved thermal coefficient is about 200 Hz.K^{-1} . Inside the vacuum system and around the cavity, a thin Al box provides additional rf shielding and thermal insulation. The vacuum system is pumped by a lateral 200 l s^{-1} ion pump which insures 10 years of pumping autonomy. Under the usual conditions, the current in the Vac-Ion pump is less than 1 mA.

Magnetic shielding is produced by five mumetal screens. The measured shielding factor is about 40 000 in the axial (vertical) direction and 400 000 in the horizontal one. The magnetic homogeneity is good enough to enable oscillation at a "C" field of about $50 \text{ } \mu\text{Gauss}$.

The hydrogen source is composed of a 100 cm^3 pyrex bulb surrounded by the rf oscillator inductance. The required dc power is about 8 Watts. The discharge is self starting and does not need regulation. The design of the controlled palladium leak is similar to that published previously². The state selector is a classical hexapole magnet with a bore diameter of 3 mm, placed at 24 cm from the storage bulb entrance ($T_B \sim 1 \text{ s}$).

The thermal regulation of the maser cavity consists of three concentric independent ovens with proportional control provided by varying the duty cycle of the current in coaxial heaters (Fig. 2). The operating temperatures of the ovens are 30°C , 42°C and 45°C . Ambient temperature changes are reduced by a factor of about 1000. The dc power consumption is about 20 Watts.

The overall electrical consumption is 100 W at 24 Vdc. The internal battery (24 Vdc, 20 Ah) provides a 3 hours minimum autonomy. The masers can be operated from 220 Vac ; in that case, the internal switching power supply simultaneously ensures battery charging and maser operation.

I. Physical Package (Fig. 1)

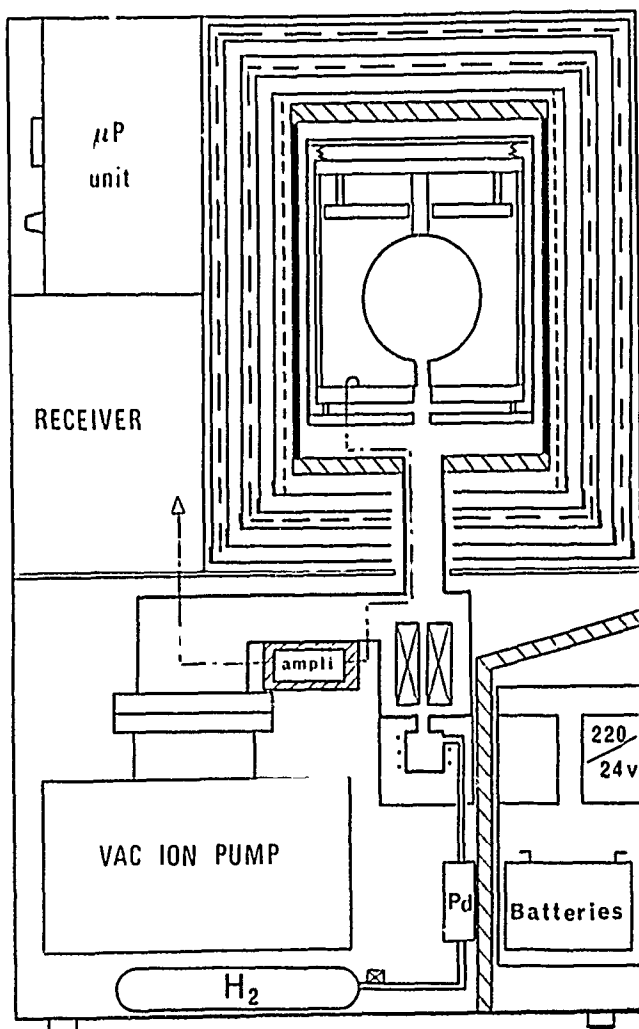


FIG.1-OVERALL PHYSICAL PACKAGE

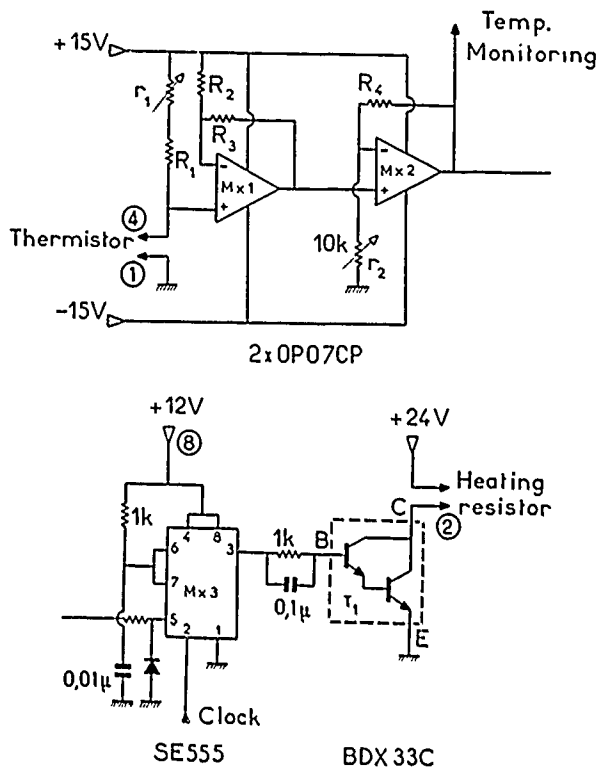


Fig.2 Thermal regulation

II. Electronics

II.1. Receiver and Phase Locked Loop

The receiver (Fig. 3) is a classical superheterodyne converter with 3 intermediate frequencies at 19.6 MHz, 405 kHz and 5.75 kHz. A digital synthesizer delivers a 575 kHz signal from which a TTL counter (1/100) provides the 5.75 kHz local oscillator. The frequency of the 5 MHz phase locked VCXO can be adjusted by means of the synthesizer in a relative range of $\pm 0.68 \cdot 10^{-7}$ with relative steps of $8 \cdot 10^{-15}$.

To obtain the best frequency stability transfer of the maser signal to the VCXO, we have minimized the random and deterministic fluctuations and optimized the phase locked loop.

To achieve the first goal, we need i) an ultra low noise and high gain 1.42 GHz amplifier : $NF = 1$ dB, $G = 24$ dB with AsGa FETs, ii) a good filtering of the image frequency at each mixer : 30 MHz bandwidth filter (Lark Engineering) at the output of the 1.42 GHz amplifier, Meirimac Image Reject Mixers for the other mixer, iii) a very good phase noise performance of the $5 \rightarrow 1440$ MHz frequency multiplier, with a small thermal phase drift : $S_{\phi} = \frac{1}{f} 10^{-13.0} + 10^{-15.8} \text{ rad}^2 \text{ Hz}^{-1}$ referred to 5 MHz and one multiplier, $\phi/2\pi f < 10^{-15}$ for $1^\circ\text{C}/10$ hours ($f_m = 1.42$ GHz). Transfer of amplitude fluctuations into phase fluctuations is minimized by simple means ³.

The second requirement leads to a need for the optimum frequency cut off f_c of the phase locked loop. In laboratory conditions, it is given by the crossing point of spectral densities of fractional frequency fluctuations S_y for the maser oscillator and for the free running 5 MHz oscillator. With the OSA 8601 quartz oscillator, and the maser oscillator under our operating

conditions ($P_{in} \sim -110$ dBm), we find $f_c = 0.5$ Hz. The operational filter's transfer function is $F(p) = (1 + \tau_1 p) / \tau_2 p$; the damping factor is adjusted to unity. Finally, the characteristics of the phase locked loop are kept constant by the use of an amplitude limiter at the rf input of the 5.75 kHz phase detector.

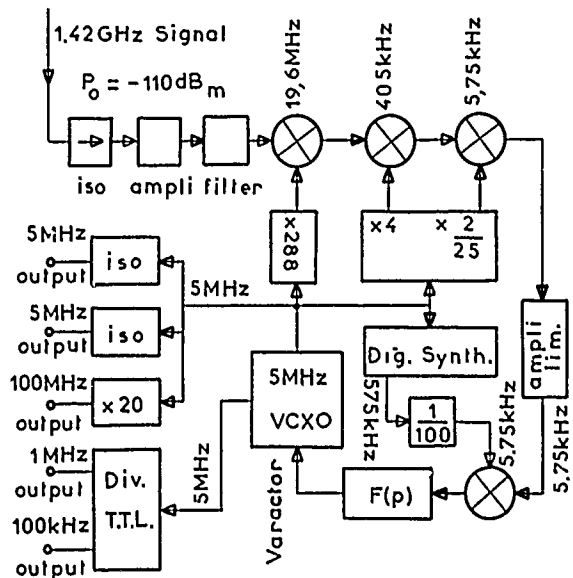


Fig.3_Receiver and P.L.L.

II.2. Microprocessor Unit

The control and the monitoring of the maser operation are made using a μ P-unit (6502).

A small keyboard allows function selection and data input. The μ P unit controls the H_2 pressure, the magnetic field value, the cavity tuning, and the frequency synthesizer in the receiver. It monitors 32 parameters of interest : supplies, local oscillators, maser signal, servo... in the receiver and PPL ; supplies, thermal regulations, ion pump current in the physical package. It provides an alarm when necessary.

The monitored parameters and controlled functions are displayed and may be printed, which allows an easy monitoring of the masers at a distance. In addition, the L.C. display indicates date and hour.

II. 3. Available Signals

II.3.1. The very good frequency stability signals are two 5 MHz signals (+ 7 dBm) and one 100 MHz signal (> + 7 dBm), whose short term frequency stability is about $3 \cdot 10^{-13} \tau^{-1}$ and medium term stability much better than 10^{-14} (Fig. 4). Isolation between the 5 MHz output is better than 95 dB under operating conditions.

II.3.2. The good frequency stability signals are 1 MHz and 100 kHz TTL signals, which may be used as an external standard for frequency counters.

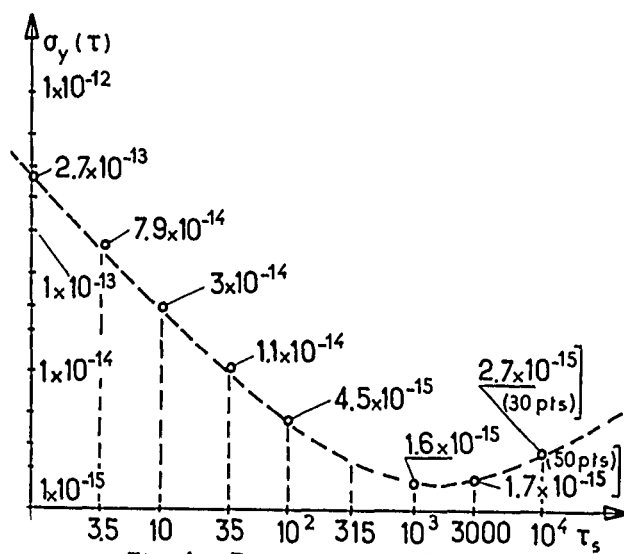


Fig.4 Frequency stability

III. Measurement results

III.1. Stability Measurements

These are made at 5 MHz with a stability measurement system^o whose noise floor is better than $10^{-13} \tau^{-1}$ at 5 MHz, for $\tau = 1$ s to 1 000 s. This system allows stability measurements for frequencies in the range 1 - 500 MHz. Operating conditions are : weak H_2 pressure (Vac-Ion pump current < 1 mA), $P_0 = -110$ dBm at the output of the 1.42 GHz cavity, $B_{int} = 1$ mG. The measurement bandwidth is 12 Hz and the maser bandwidth considered is 1 Hz. The number of points used for each value of σ_y is > 100 if not indicated between brackets (Fig. 4).

The frequency stability σ_y (Fig. 4) is very good between $\tau = 1$ s and $\tau = 10^4$ s, which was the requirement for V.L.B.I. applications. We did not measure stability for $\tau > 10^4$ s, but one can estimated that σ_y is better than 10^{-14} for $\tau = 10^5$ s.

Stability measurements made at 100 MHz give the same results as previously indicated in Fig. 4 for $\tau = 1$ s to 100 s (no longer measurements made).

III.2. Systematic Effects

III.2.1. Thermal fluctuations. We have measured thermal effects with a temperature step of 3°C . For the 3 masers, the thermal sensitivity is $|\frac{\Delta f}{f \Delta T}| \sim 1.10^{-13} \text{ } ^\circ\text{C}^{-1}$. The signs are not the same for 3 masers, due to the differences between the cavity thermal compensations. This thermal sensitivity requires that the masers be located in a room having a good thermal regulation.

III.2.2. Magnetic fluctuations. We have measured the magnetic sensitivity of the 3 Masers in the axial (vertical) direction, which is the most sensitive. Helmholtz coils surround one maser, while a second maser is used as a reference. The internal static magnetic field is 1 mG.

The measured effects are not linear, so we distinguish two types of magnetic fluctuations :

- for usual weak magnetic fluctuations ($\Delta B_{ext} < 50$ mG), the measured value is approximately the same for the 3 masers, and it is in good agreement with the predicted value :

$$\frac{\Delta f}{f \Delta B_{ext}} = 1.10^{-13} \text{ G}^{-1}$$

- for an external strong magnetic fluctuation ($\Delta B = \pm 1$ Gauss), the measured values are not the same for the 3 masers. The mean behaviour is

$$|\frac{\Delta f}{f}| \sim 3 \cdot 10^{-13} \text{ for } |\Delta B| = 1 \text{ Gauss}$$

Conclusion

The 3 masers manufactured at L.H.A. (Orsay, France) are to be used as frequency references in V.L.B.I. experiments. Their stability in short and medium term is quite convenient for that use. In addition, thermal and magnetic sensitivities are small enough to allow a frequency stability of a few 10^{-15} in medium term.

Acknowledgements

The realization of the three masers was supported by the Direction des Recherches et Etudes Techniques (D.R.E.T.), the Institut National des Sciences de l'Univers (I.N.S.U.) and the Centre National d'Etudes Spatiales (C.N.E.S.).

References

1. P. Petit, J. Viennet, R. Barillet, M. Desaintfuscien and C. Audoin, "Development of Hydrogen Masers as Frequency Standards at the Laboratoire de l'Horloge Atomique". *Metrologia* **10**, p. 61-67, 1974
2. J. Viennet, P. Petit and C. Audoin, "Régulateur de débit d'hydrogène à réponse rapide". *Journal of Physics E : Scientific Instruments* 1973, Vol. 6, p. 257-261
3. R. Barillet and C. Audoin, "Electronic Circuits Showing a Level Independent Phase Shift for Hydrogen-Maser Work Applications". *IEEE Trans. on Instr. and Meas.* **IM-25**, n° 4, Dec. 1976, p. 465-468
4. R. Barillet, Thèse de Docteur Ingénieur (Février 1977, Orsay). "Accord de la cavité résonnante d'une horloge atomique à hydrogène par la méthode de la phase".
5. C. Audoin, "Frequency Metrology". *Metrology and Fundamental Constants* (North Holland Publ. Co.) 1980. p. 171
6. P. Petit, R. Barillet, J. Viennet and C. Audoin, "Masers à hydrogène actifs et compacts". *Congrès International de Chronométrie*, Besançon, 1984

A HYDROGEN MASER AT TEMPERATURES BELOW 1K

Robert F.C. Vessot and Edward M. Mattison
Harvard-Smithsonian Center for Astrophysics
Cambridge, MA 02138

and

Ronald L. Walsworth, Jr., Isaac E. Silvera, H.P. Godfried and C.C. Agosta
Lyman Laboratory of Physics
Harvard University, Cambridge, MA 02138

INTRODUCTION

Since the invention of the atomic hydrogen maser in 1960 by Kleppner, Goldenberg, and Ramsey[1] there has been an increasing level of activity in the search for improvement of the maser's hydrogen storage technique. The ability to keep a constantly replenished quantity of hydrogen atoms localized for extended times in a region where they are exposed to a uniform, constant-phase rf magnetic field oscillating at the hyperfine resonance frequency results in a narrow oscillator linewidth, which is the principal reason for the success of the hydrogen maser as a very high stability oscillator. In the maser a beam of hydrogen atoms in the two upper hyperfine energy levels ($F=1$, $m_F=0$ and $m_F=1$) is state selected by a strong, inhomogeneous magnetic field. The state selected beam is directed into the storage region, where atoms in the $F=1$, $m_F=0$ state undergo stimulated transitions to the $F=0$, $m_F=0$ state, coherently contributing energy to the rf magnetic field. The stored atoms move at thermal velocities associated with the temperature of the boundaries of the storage region (which for historical reasons is often referred to as the storage "bulb"). The linewidth Δf of the released energy depends on the length of time, Δt , the atoms are exposed to the rf field in an unperturbed manner, as described by the Heisenberg uncertainty principle, $\Delta E \Delta t \sim h$ (or $\Delta f \Delta t \sim 1$). Here $E=hf$ is the energy of the hyperfine separation, h is Planck's constant and f is the frequency of the transition.

The key to the whole process is the ability of the storage region walls to reflect the incoming atoms without excessively disturbing the phase of their oscillating dipole moments. Usually the dimensions of the bulb are commensurate with half the 21 cm wavelength of oscillation. At room temperature a typical atom collides with the wall $3 \cdot 10^4$ times per second. The velocity of impact is about 3.6 km/sec. That any surface material can make possible such collisions without disrupting the phase of the atoms oscillating dipole moment is indeed remarkable.

The first wall coatings used in room temperature masers were of dimethyl dichlorosilane, and later Teflon was (and still is) used. The possibility of low-temperature operation of hydrogen masers arose in 1977; about this time the suggestion was made by I.I. Shapiro[2] that considerable advantage to hydrogen masers might result in such low temperature operation if suitable wall surfaces could be found. Crampton, Phillips and Kleppner[3] described the advantages resulting from low thermal noise, and the mechanical, thermal, and magnetic stability of materials. The situation was very tempting indeed, if only the wall question could be resolved. Attempts to use frozen-in-place coatings of CF_4 at about 35K[4] and neon at about 6K[5] have been made, but these do not appear to have much promise for highly stable oscillators.

In order to understand how low temperature operation can improve the frequency stability of the hydrogen maser, we examine the Allan variance[6], σ , of the maser's frequency fluctuations. We express the

Allan variance, which is a function of the maser's power output, oscillation line Q, receiver noise figure and bandwidth, and temperature,[7,8] in terms of the maser's oscillation parameter q.[9]

Following Vessot et.al.[10] and Berlinsky and Hardy[11], we recognize two major contributions to instability that result from thermal noise:

1. Noise within the bandwidth of the receiver system, including the excess noise of the first amplifier stages.
2. The thermal noise within the oscillation linewidth, which defines the Q of the atomic system. This noise is a fundamental performance limit.

The observed frequency pulling, owing to mistuning of the cavity resonator, can also be included in a statistical sense by postulating that the spectrum, $S_y(f)$, of the cavity offset frequency follows some combination of $1/f$ (flicker of frequency) and $1/f^2$ (random walk of frequency) distribution[12] and can, therefore, be characterized by an Allan variance $\sigma_c(\tau)$.

The expression describing the stability is given by

$$\sigma^2(\tau) = kT \left(\frac{FB}{P_0 \omega_0^2 r^2} + \frac{1}{2Q_L^2 P} \frac{1}{r} \right) + \sigma_c^2(\tau) \frac{Q_c^2}{Q_L^2} \quad (1)$$

Here k is Boltzmann's constant, T is the physical temperature, assumed to be uniform throughout the system, F is the noise figure of the receiver and B its noise bandwidth, $\omega_0 = 2\pi f_0$ is the angular frequency of oscillation, P_0 is the power delivered to the receiver, P is the power delivered to the cavity by the atoms, and Q_c is the loaded Q of the cavity resonator. The oscillating line Q of the atomic resonance is given by $Q_L = \frac{\omega}{2\gamma_{2T}}$, where γ_{2T} is the total transverse (dephasing) relaxation rate, which will be discussed later.

We can express Eq. (1) in terms of the internal parameters of the maser as defined by Kleppner et.al.[9] The relationship between the power, P , radiated by the hydrogen atoms, and the flux, I , of atoms in the $F=1$, $m_F=0$ state entering the storage volume, can be expressed in terms of the maser's oscillation parameter, q , defined by

$$q = \frac{h}{8\pi\mu_0} \frac{I_{\text{total}}}{I} \left(1 + \frac{\gamma_w}{\gamma_b} \right) \frac{V_c}{\eta V_b} \frac{\sigma_{se}(T) \bar{v}_r(T)}{Q_c}$$

Fig. 1, from Kleppner et.al., shows how the quantity $\frac{P}{P_c} = -2q^2 \left(\frac{I}{I_{\text{th}}} \right)^2 + (1-3q) \frac{I}{I_{\text{th}}} - 1$ behaves in terms

of the parameter q . Here $P_c = \frac{\omega h^2 V_c (\gamma_b + \gamma_w)^2}{8\pi\mu_0^2 Q_c \eta}$, $I_{\text{th}} = 2P_c/h\omega$, γ_w is the dephasing relaxation rate of atoms from wall collisions, and γ_b is the loss rate of atoms from the storage bulb. We assume here that γ_b

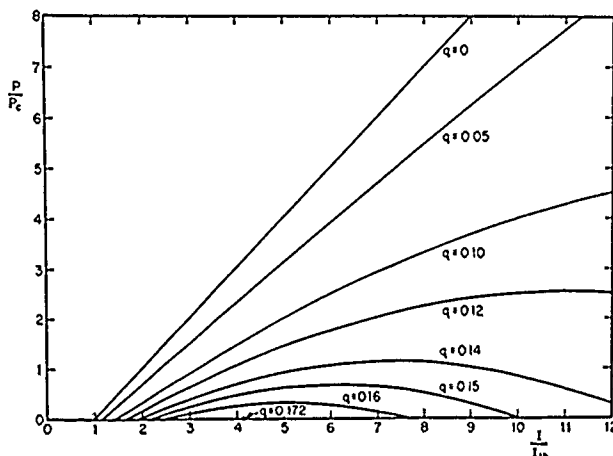


Fig. 1. P/P_c versus I/I_{th} for different values of the parameter q .

is due to all mechanisms causing loss of atoms, including recombination on the storage bulb walls as well as escape through the bulb's entrance aperture. We assume that the rate of loss of polarization and the rate of dephasing owing to wall collisions are equal. Atoms in the storage bulb will also suffer dephasing resulting from spin-exchange collisions. The rate $\gamma_{se} = n\bar{v}_r\sigma_{se}$ of such collisions is governed by the atomic density, n , the spin-exchange cross section, $\sigma_{se}(T)$, and the relative atomic velocity, $\bar{v}_r(T) = 4(kT/\pi m)^{1/2}$. The total transverse relaxation rate is given by $\gamma_{2T} = \gamma_b + \gamma_w + \gamma_{se}$. The quantity V_c/V_b is the ratio of the resonant cavity volume to the storage volume. $\eta = \langle H_z \rangle_b^2 / \langle H^2 \rangle_c$ is the filling factor representing the ratio of the rf magnetic field engaged in the transition, to the total rf magnetic field in the resonant cavity. I_{tot} is the total flux of atoms entering the bulb (which, under the usual state-selection conditions, is twice the flux I in the desired $F=1$, $m_F=0$ state) and μ_0 is the Bohr magneton.

We note that the quadratic behavior of power with beam flux results from spin-exchange quenching. The parameter q expresses the specific conditions inherent in the maser's design and operating temperatures; for oscillation to occur, q must be less than 0.172.

We can recast Eq. (1) in terms of q and the beam flux I by substituting the appropriate relationship for Q_L and output power. The line Q can be expressed by:

$$Q_L \equiv \frac{\omega}{2\gamma_{2T}} = \frac{\omega}{2(\gamma_b + \gamma_w)} \left(1 + q \frac{I}{I_{th}} \right) \quad (2)$$

The three noise contributions, which include the assumed spectral behavior of the resonator frequency variations, $\sigma_c(r)$, then appear as follows:

$$\sigma^2(r) = \frac{\frac{8\pi\mu_0^2 F k T B \eta}{\omega^3 h^2} \frac{Q_{ext}}{V_c (\gamma_b + \gamma_w)^2}}{-1 + (1 - 3q) \frac{I}{I_{th}} - 2q^2 \left(\frac{I}{I_{th}} \right)^2} r^{-2} + \frac{\frac{16\pi}{\omega^3} \frac{k T \mu_0^2 \eta}{h^2 V_c} Q_c \left[1 + q \left(\frac{I}{I_{th}} \right) \right]^2}{-1 + (1 - 3q) \frac{I}{I_{th}} - 2q^2 \left(\frac{I}{I_{th}} \right)^2} r^{-1}$$

$$+ \frac{4Q_c^2}{\omega^2} (\gamma_w + \gamma_b)^2 \left[1 + q \left(\frac{I}{I_{th}} \right) \right]^2 \sigma_c^2(r) \quad (3)$$

Here $\frac{1}{Q_c} = \frac{1}{Q_0} + \frac{1}{Q_{ext}}$, where Q_0 is the unloaded Q of the resonator and Q_{ext} represents the loading due to external systems coupled to the cavity. This load should be as non-reactive as possible so as to avoid excessive mistuning of the loaded system by external variations.

In a typical maser q is about 0.07 and the relationship between actually measured stability and the stability predicted from operating parameters has been accurately tested.[10]

The benefits of low temperature operation are vividly apparent in the makeup of the parameter q , which contains the factor $\sigma\bar{v}_r$. At very low temperature this factor decreases by a factor of about 10^3 from the room temperature value. For a given maser configuration and assuming γ_w is negligible, we see that we can reduce Q_c very substantially and still, under very cold conditions, maintain the same oscillating condition represented by q . Cavity Q reduction is best done by over-coupling the cavity resonator to the external load, thus delivering considerably more power to the load. A further advantage results from the reduced cavity pulling (provided the load is non-reactive and stable). The increased power substantially improves the stability. When the reduction in kT is included in the signal-to-noise ratio, we anticipate even further improvement in stability.

The line Q is also affected by reducing the temperature. Q_L is defined by $Q_L = \omega/2\gamma_{2T}$, and $\gamma_{2T} = \gamma_b + \gamma_w + \gamma_{se}$. We note that all three terms of γ_{2T} depend on velocity and hence on T as follows:

$\gamma_b = \frac{\bar{v}A_0}{4\kappa V_b}$ is the escape rate term, where A_0/κ is the exit aperture area divided by the kappa factor[13] of the bulb collimator.

$\gamma_w = \bar{v}p/\lambda$ is the wall relaxation term. Here λ is the mean-free distance of the atoms in the storage volume and p is the probability per collision of the atom losing phase with respect to the rf magnetic field.

$\gamma_{se} = n\bar{v}\sigma$ is the spin-exchange term, which has already been discussed and is included in terms of q and I in Eq. (2).

We see that simply because the atom's velocity is reduced and because the spin exchange cross section is much smaller, lowering the temperature can improve the atomic Q provided the wall collision dephasing term is very small compared to the other terms.

LIQUID HELIUM WALL COATINGS

In 1979 Silvera and Walraven,[14] at the University of Amsterdam, demonstrated long-term stabilization of a moderate density of atomic hydrogen at low temperature. A critical requirement was coverage of all confining walls with a film of liquid ^4He (which is superfluid). This inert wall coverage provides an almost ideal surface for a hydrogen maser. Subsequently, Greytak, Kleppner and his colleagues at M.I.T. and the Amsterdam group succeeded in obtaining magnetically confined densities as high as $10^{17}/\text{cm}^3$ in the neighborhood of 0.3K.[15,16] Parallel work by at the University of British Columbia by Berlinsky and

Hardy[17] with closed cells at very low magnetic fields demonstrated the feasibility of making transitions between the magnetic resonance hyperfine levels of the hydrogen atoms. These experiments provided a great deal of information on the interactions of atomic hydrogen on the fluid helium surfaces and in the helium gas over the surfaces. The prospects for applying helium wall surfaces have been described by Hardy and Morrow[18] and by Berlinsky and Hardy[11], who showed that the wall shift and pressure shift are both negative, reducing the transition frequency. They found that the combined wall and pressure shift has a maximum (is least negative) around 0.5K so that operation in this region can be made stable with respect to temperature variations.

Successful oscillation of hydrogen masers at very low temperatures has been reported by both the M.I.T.[19] and the University of British Columbia groups.[20] Parallel efforts to design, build, and operate a low temperature maser at Harvard have been rewarded by successful oscillation.[21]

THE HARVARD-SMITHSONIAN CRYOGENIC HYDROGEN MASER

We have designed, built, and demonstrated operation of a cryogenic maser (CHM) operating in the temperature range of 350 to 600 millikelvin. Our maser design is similar to that of a conventional room temperature maser, except that the atomic hydrogen beam is cooled to low temperature, and the resonator consists of a dielectrically-loaded cylindrical cavity made of sapphire.[22] A novel feature is a continuously flowing superfluid helium film that covers the maser storage volume walls to prevent recombination of the hydrogen atoms and preserves the phase of their oscillating dipole moment. This film is eventually adsorbed by a large sorption pump operating at 6 K. The CHM has been predicted by Berlinsky and Hardy[11] to have frequency stability in the range 10^{-18} for averaging times of about 10^3 seconds. Although room temperature hydrogen masers have been operated with stabilities[23] of a few parts in 10^{16} , their accuracy is limited to several parts in 10^{13} by the irreproducible wall shift of Teflon surfaces. Preliminary performance results of our maser, which has been designed to yield the high predicted stability, are reported here.

Our maser is shown in Fig. 2. Atomic hydrogen in all four hyperfine states flows out of a room temperature microwave discharge into the cryostat through a Teflon-lined tube and is thermalized by a 3 mm-diameter source aperture at 10K, as described by Walraven and Silvera.[24] The cold beam then enters a 1.3 cm long hexapole magnet that focusses atoms in the ($r=1$, $m_F=0$ and $m_F=1$) hyperfine states into the aperture of the cylindrical dielectrically loaded microwave cavity. An axial beam stop at the exit of the hexapole magnet prevents undeflected atoms from entering the cavity. The cavity is mounted in good thermal contact within a large copper isothermal chamber that is suspended from the mixing chamber of a dilution refrigerator and can be cooled down to about 300 mK. In order to attenuate the earth's magnetic field and achieve high frequency stability the cavity is surrounded by a four-layer nested shield made of Cryo-perm, a material with high magnetic susceptibility at low temperatures. These shields operate at 4.2K and are thermally isolated from the cavity. A uniform axial magnetic field of a few milligauss is produced in the storage volume by a solenoid made up of three separately energized coils. The solenoid is located outside the copper cylinder and within the innermost magnetic shield. A two loop transverse coil (not shown) is attached to the wall of

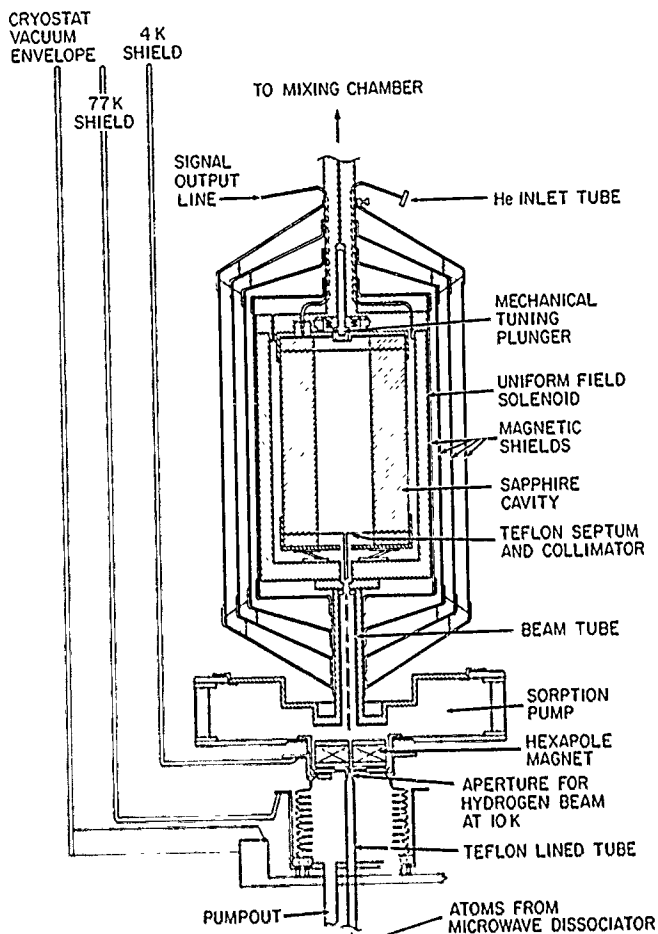


Fig. 2. Schematic diagram of cryogenic hydrogen maser.

the microwave cavity. It is used to measure the cavity field by inducing audio frequency $\Delta m_F = \pm 1$ transitions that are detected by quenching of the maser oscillation. Another coil, mounted outside the beam tube between the hexapole and the cavity aperture, provides a small static "holding field" to prevent depolarization of the state-selected beam due to Majorana transitions.

The resonant cavity consists of three pieces of c-cut single-crystal sapphire. The central cylindrical section has an axial hole through its length that forms the hydrogen storage volume. The cylinder is capped by two annular sapphire discs. The interior of the cylinder is coated with Teflon for tests at higher temperatures[23] and the discs are separated from the cylinder by Teflon septa. The cavity's electrically conducting surfaces are cup-shaped copper plates at the top and bottom, and thin copper shim stock tightly wrapped around the cylinder. A Teflon tube (4 mm id and 27 mm long) passes through the lower septum to admit hydrogen atoms into the storage volume. By confining the atoms to a region where the microwave field is uniform, rather than allowing them to occupy the ends of the cavity, the septa improve the filling factor. The dimensions of the Teflon tubes determine the storage time of the atoms in the cavity, which is calculated to be 7.5s at 0.5K. The cavity frequency can be mechanically tuned over a 600 kHz range with an externally operated cylindrical plunger. This allows us both to tune the cavity to resonance and to measure the atomic line width, or line Q, by detuning the cavity from resonance and measuring the ensuing maser frequency shift. The cavity is weakly coupled to a 50 ohm output coaxial cable by a pickup loop near the end of the cavity. The cavity Q is about 20,000 at room temperature and increases to 65,000 at low temperature. Due to the temperature dependence of the

sapphire's dielectric coefficient, the cavity frequency increases by about 12 MHz in cooling from room temperature to 4K. Preliminary measurements enabled us to size the sapphire so that at about 0.5K the cavity frequency can be tuned to the hydrogen maser transition frequency of 1,420,405,751 Hz.

In order to prevent recombination of hydrogen at subkelvin temperatures, all surfaces must be covered by a film of liquid helium.[14] Since the ^4He is superfluid at the CHM operating temperatures, it flows to cover all surfaces, but in particular it is driven to warmer regions. Helium from a room temperature source of variable pressure is introduced into the cavity via a capillary tube. Without proper measures the liquid helium would flow out of the cavity aperture and up the outer walls to warmer regions where it would vaporize, destroy the vacuum, and warm the dilution refrigerator to high temperatures. We have built a giant sorption pump between the beam tube and the hexapole magnet. The pump chamber is filled with activated charcoal that is thermally linked to the pump's bottom plate at 6K. The top plate operates at the maser temperature and is thermally isolated from the bottom plate by a 25 μm thick Kapton wall. The two plates are structurally strengthened by six tubes of G-10 fiberglass-reinforced epoxy. The helium film vaporizes at the Kapton walls and, due to baffles, the vapor is cryopumped before diffusing back to the central region, where it would attenuate the hydrogen beam.

At higher temperatures, of order 550 mK, the vapor pressure of a saturated film of He is high enough that it ordinarily would attenuate the beam in the beam tube region, reducing the population of $F=1$, $m_F=0$ state atoms below the threshold for oscillation. To prevent this we have built a novel "dynamic" He pump. Since the vapor pressure of very thin films of He is substantially lower than the saturated value, we have placed a large number of concentric sheets of copper foil in the beam tube areas so that the helium film that flows out of the cavity aperture is subdivided into a large number of thin undersaturated films with a low vapor pressure that flow in parallel along the sheets. In addition, the cryopumping copper surfaces also adsorb helium that flows out of the aperture in the vapor phase. Our current tests of the performance of the dynamic pump are inconclusive.

MEASUREMENTS

We have observed maser oscillation of our CHM from approximately 350 to 575 mK at power levels up to 5×10^{-13} watts. Due to technical cryogenic problems we were unable to make measurements for periods longer than about three hours; more important, we were limited to helium flow rates that we estimate are lower than those required to achieve saturated films. (We assume the film thickness to be essentially proportional to flow rate until the film is saturated.) As a consequence, in this article we concentrate on the relationship between the film thickness and both the maser frequency (i.e. the wall shift) and the line width or line Q, which is important for the frequency stability of the maser. The predicted frequency shifts from helium surface and bulk gas effects are shown in Fig. 3, along with the mean free path of hydrogen in the helium vapor.

The flow rate is determined by the pressure, p , in a room temperature helium reservoir. Since the flow should be proportional to p^2 , so should the film thickness. In Fig. 4 we show the wall shift Δf_w (that is, the shift of the maser frequency from the value it would have for free atoms in the absence of walls) as a function of p^2 at a constant temperature of 0.493 K. At this temperature the wall shift for saturated

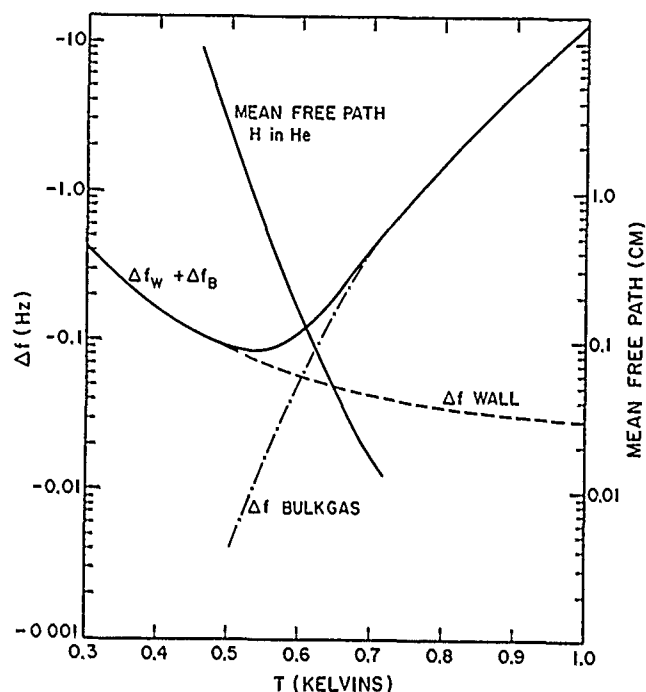


Fig. 3. Predicted frequency shift and mean free path for hydrogen in cryogenic maser.

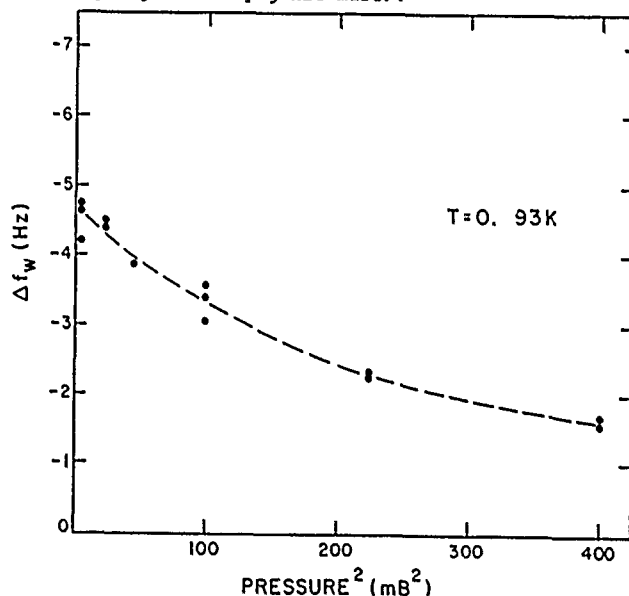


Fig. 4. Measured wall shift in cryogenic maser.

surfaces is expected to be on the order of 0.1 Hz. At very small values of flow the wall shift is substantial and decreases toward the saturated value as flow or film thickness increases. This is reasonable, as for thinner films the hydrogen atoms see the more polarizable substrate[25] and we would expect the hyperfine frequency to decrease.

Fig. 5 shows the wall shift as a function of temperature for fixed He flow rates. For a saturated film this curve is expected to be bowl shaped, the shift decreasing as temperature increases to about 0.6K and then increasing at higher temperatures due to the pressure shift. Our data, which are for $T < 0.6\text{K}$, are consistent with this behavior. In Fig. 5 we also show the line Q as a function of temperature for fixed He flow rates. If the line Q is limited only by the cavity escape time (7.5s), then we expect $Q_L = 3 \times 10^{10}$. The measured values of Q_L are much smaller than the calculated value, but increase rapidly with

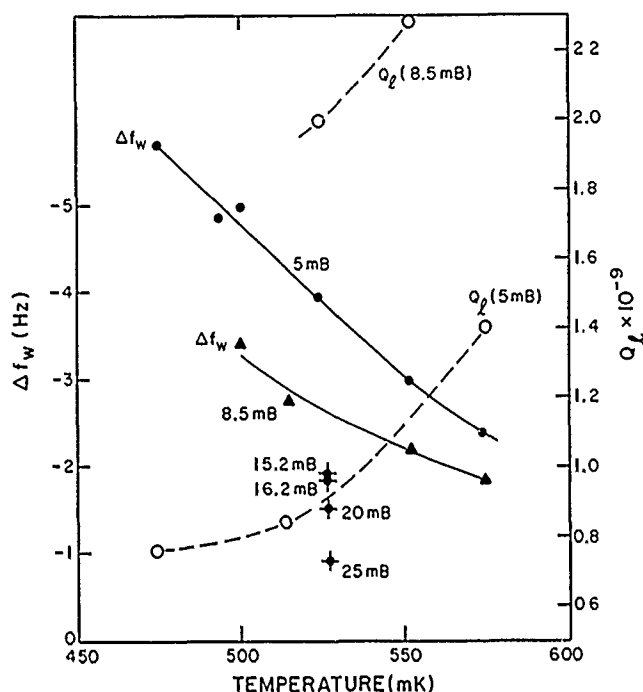


Fig. 5. Measured wall shift and line Q as functions of temperature.

temperature and, more important, with He film thickness. In order to characterize the CHM as an oscillator of extreme frequency stability, it will be necessary to extend these measurements to saturated values of the helium film.

CONCLUSION

The stability we expect to observe, assuming the performance indicated by Eq. (3), is shown in Fig. 6 as the lower dotted curve. The currently available stability of the SAO VLG-11 masers is shown in a solid line, as is the stability of the Hewlett-Packard mercury ion standard.[26] The projected stability[12] of room temperature hydrogen masers, operating with an improved atomic state selection system, is shown as a dotted curve that just touches the 1×10^{-16} level at 3×10^4 sec.

We realize that there are many systematic effects capable of affecting the maser's output frequency that are very large compared to the statistically represented performance. The traditional sources of instability such as cavity-pulling and magnetic variations seem to be well within our ability to control owing to the low cavity Q , high line Q , and use of superconducting magnetic shields. The atomic hydrogen wall and bulk gas frequency shifts shown in Fig. 3, demand control of temperature at the 0.53 K nullpoint within 20 μ K if 10^{-18} stability is to be achieved. This, too, is within the present state of technology.[27]

So far, as we have yet to achieve a fully saturated helium film in the storage volume, we are uncertain as to the required hydrogen flow stability, let alone the stability of the flowing film itself or its dynamics as seen from the perspective of the incoming hydrogen atoms. We expect that ongoing measurements will help to resolve some of these questions and lead us to new information about hydrogen-helium surface interactions and to the development of better clocks.

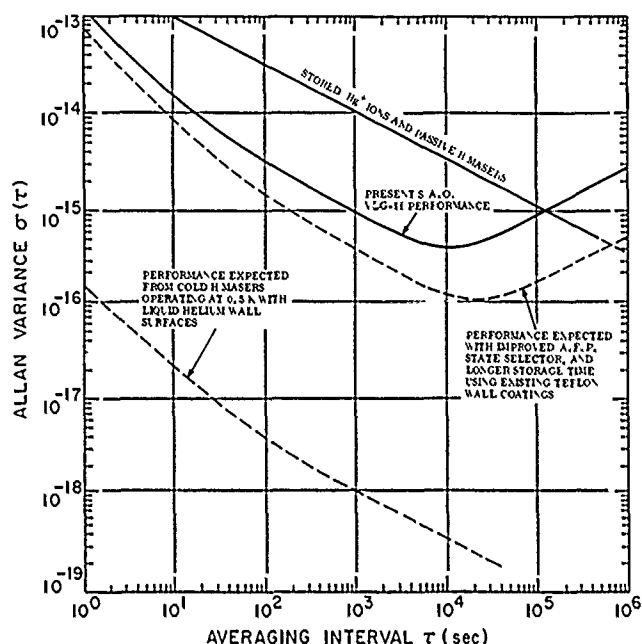


Fig. 6. Predicted frequency stability for cryogenic hydrogen maser compared with other high-performance clocks.

ACKNOWLEDGEMENTS

We would like to thank Mr. Pierre Crevoisierat for his expert craftsmanship in building the maser. We also thank Mr. Donald Gravelline, Mr. Richard Nicoll and Mrs. Shen-Wei for their valuable help in preparing and testing the sapphire resonator and the hexapole magnet. This work was done under the Smithsonian Institution's Scholarly Studies and Research Opportunities Fund and by the Department of Energy, Grant DE-FC02-85ER45190. One of us, H.P.G., was supported as an I.B.M. post-doctoral fellow.

REFERENCES

1. Kleppner, D., H.M. Goldenberg, and N.F. Ramsey, Phys. Rev. **126**, 603 (1982)
2. Shapiro, I.I., Private Communication (1977).
3. Crampton, S.B., W.D. Phillips and D. Kleppner, Bull. Am. Phys. Soc. **23**, 86 (1978)
4. Vessot, R.F.C., E.M. Mattison, W.J. Klepczynski, I.F. Silvera, H.P. Godfried and R.L. Walsworth, Jr. Proc. 16th Annual Precise Time and Time Interval Meeting, 413 (1985).
5. Crampton, S.B., K.M. Jones, G. Nunes, and S.P. Souza, Proc. 16th Precise Time and Time Interval Meeting, 339 (1984).
6. Allan, D.W., Proc. I.E.E.E., **54**, 221 (1966).
7. Vessot, R.F.C., Quantum Electronics III, P. Grivet and N. Bloembergen, eds., Columbia Univ. Press, New York, 409 (1964).
8. Cutler, L.S. and C.L. Searle, Proc. I.E.E.E. **54**, 136 (1966).
9. Kleppner, D., H.C. Berg, S.B. Crampton, N.F. Ramsey, R.F.C. Vessot, H.E. Peters and J. Vanier, Phys. Rev. **138**, 972 (1965).

10. Vessot, R.F.C., M.W. Levine, and E.M. Mattison, Proc. 9th Annl. Precise Time and Time Interval Meeting, 547 (1977).
11. Berlinsky, A.J., W.N. Hardy, Proc. 13th Annual Precise Time and Time Interval Meeting, 547 (1981).
12. Mattison, E.M., R.F.C. Vessot and Shen-Wei, Proc. 40th Symposium on Frequency Control (1986). In press.
13. Clausing, P., Physica, 2, 65 (1929)
14. Silvera, F. and J.T.M. Walraven, Phys. Rev. Lett. 44, 164 (1980).
15. Cline, R.W., T.J. Greytak, D. Kleppner and D.A. Smith, J. de Physique 41 07-151 (1980).
16. Walraven, J.T.M. Walraven, I.F. Silvera, and A.P.M. Matthey, Phys. Rev. Lett. 45, 449 (1980).
17. Morrow, M., R. Jochemsen, A.J. Berlinsky, and W.N. Hardy, Phys. Rev. Lett. 46, 145 (1981). Erratum 47, 455 (1981).
18. Hardy, W.N., M. Morrow, J. de Physique 42, C8-171 (1981).
19. Hess, H.F., G.P. Kochanski, J.M. Doyle, T.J. Greytak and D. Kleppner. Preprint.
20. Hurlimann, M.D., W.N. Hardy, A.J. Berlinsky and R.W. Cline, Preprint.
21. Walsworth, R.L., I.F. Silvera, H.P. Godfried, C.C. Agosta, R.F.C. Vessot and E.M. Mattison, Phys. Rev. Rapid Communication, in press.
22. Folen, V.J., W.G. Mausch, J.D. White, C.A. Bartholomew, and A.J. Frank, N.R.L. memorandum report 487D, U.S. Naval Research Laboratory, Washington, DC (1982)
23. Vessot, R.F.C., E.M. Mattison, W.J. Klepczynski, I.F. Silvera, H.P. Godfried, and R.L. Walsworth, Jr. Proc. 17th Ann. Precise Time and Time Interval Meeting, 413 (1985).
24. Walraven, J.T.M., and I.F. Silvera, Rev. Sci. Instr. 53, 1167 (1982).
25. Godfried, H., E. Eliel, J. Brisson, J. Gillaspay, C. Mallardeau, and I.F. Silvera, Phys. Rev. Lett. 55, 131 (1985).
26. Cutler, L.S., R.P. Gifford, and M.D. McGuire, Proc. 37th Symposium on Frequency Control, 32 (1983).
27. Lipa, J.A., Int. Conf. Low Temp. Phys., Physica 107 B+C, 343 (1981).

A Microwave Pumped Cryogenic Hydrogen Maser

W.N. Hardy, M.D. Hürlimann, R.W. Cline, and A.J. Berlinsky

Department of Physics, University of British Columbia,
Vancouver, B.C. CANADA V6T 2A6

ABSTRACT: A cryogenic hydrogen maser of novel design with potential for extremely high stability has been developed and successfully operated over the temperature range 0.25 to 0.7 K. The maser employs liquid helium coated surfaces and operates in a closed cycle where the atoms circulate back and forth between a microwave-pumped state selector and the storage bulb. For measuring times between 1 and 300 seconds the maser showed no measurable frequency fluctuations at the level of 6×10^{-13} , the limit of our local frequency reference.

There are a number of advantages to operating an atomic hydrogen maser at very low temperatures. They include lower intrinsic noise, the use of cooled electronics to minimize added receiver noise, and the almost negligible sensitivity of physical dimensions to temperature changes. In addition, the spin-exchange broadening constant $\sigma_{ex}v$ is greatly reduced at low temperatures, being $10^{-12} \text{ cm}^3/\text{s}$ at 1 K compared to $10^{-9} \text{ cm}^3/\text{s}$ at room temperature. This means that for a given linewidth of the masing transition, one can work at much higher state selected flux and thus at higher output power. Around 1980 workers in the field of spin-polarized hydrogen demonstrated that a thin film of liquid helium was a very good, non-reactive wall coating for confining H at low temperatures. In particular, our own work on zero field hyperfine resonance showed the existence of a temperature, $T_{\min} \approx 0.5 \text{ K}$, at which the hyperfine frequency shift of an atom confined to a helium coated bulb went through a minimum¹. This is a result of the combined effect of the He vapour pressure shift and the He-coated wall shift. T_{\min} provides a relatively stable operating point for a Cryogenic Hydrogen Maser (CHM). We have predicted on the basis of a theoretical analysis², that such a maser could achieve frequency stabilities orders of magnitude better than conventional hydrogen masers which are presently the most stable frequency sources for intermediate measuring times $1 \text{ s} \leq \tau \leq 10^6 \text{ s}$ ³. For optimum parameters a CHM would have a predicted frequency stability (Allan variance) of $\Delta f/f = 2 \times 10^{-18}$ for a measuring time of 10^3 s , to be compared to the best measured stability of a conventional room temperature maser³ of 6×10^{-16} .

A simplified and partly schematic diagram of the CHM is shown in Fig.1. Atoms from a low temperature source (not shown) travel down the waveguide (A) and enter a 40 GHz microwave cavity (C) which sits at the center of a 14 kG superconducting magnet (B). Microwaves at 40 GHz are used to pump the b to c hyperfine transition (see Fig. 2), and atoms in the c-state are ejected from the high field region, down the atom tube, Fig.1 (E), and through the orifice (F) into the Pyrex maser bulb (H) which sits inside a split ring resonator⁴ (G) tuned to 1420 MHz. Atoms in the maser cavity precess and radiate at the 1420 MHz hyperfine frequency f_{ac} . Atoms which emerge from the maser bulb in the a-state are drawn back toward the state selector (B,C), passing by a "relaxing foil" in region (D). The foil induces rapid transitions between the a and b hyperfine states⁵ ($f_{ab} \approx 1 \text{ GHz}$). b-atoms which diffuse back into the state selector microwave cavity (C) can be pumped back into the c-state, thus completing the cycle. The small tip at the end of the maser bulb (H) is a reservoir for liquid ^4He . The fact that H binds very weakly to liquid ^4He surfaces greatly suppresses the recombination rate for $\text{H} + \text{H} \rightarrow \text{H}_2$ due to surfaces. As a result, H atoms can be recirculated back and forth between the state selector and maser bulb many times

because their recombination lifetime, for the H atom densities used in the CHM, is very long. The CHM was first operated successfully in our lab in April 1986, and a brief report on its operation has been submitted elsewhere.⁶

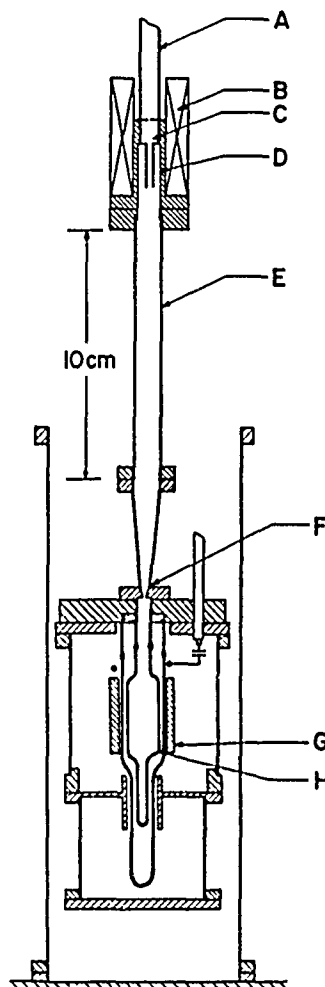


FIG. 1. Simplified and partly schematic diagram of the cryogenic hydrogen maser drawn to scale. The various parts and the operation of the maser are described in the text.

Masing occurs because the state selector produces a non-zero polarization density $\delta_{ac}^0 = n_c^0 - n_a^0$ outside the orifice of the maser bulb. This polarization density, together with the bulb volume, $V_B = 5.15 \text{ cm}^3$, and the bulb holding time, $T_B = 0.6 \text{ s}$, define an effective flux

$$I = \delta_{ac}^0 V_B / T_B$$

of polarized atoms into the bulb. The power radiated by the atoms is

$$P = \frac{\hbar \omega}{2} (I - I_0 T_B^2 / T_1 T_2)$$

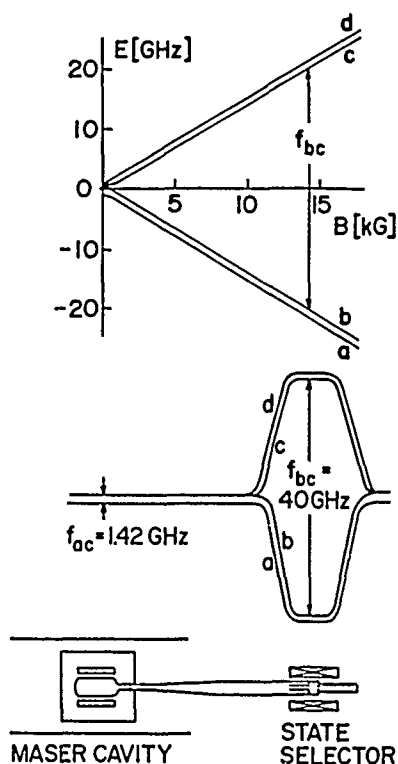


FIG. 2. Breit-Rabi diagram for H showing labeling of states, along with a schematic indication of the energy splittings as a function of position in the recirculating cryogenic hydrogen maser.

where $1/T_1 = 1/T_D + \sigma_{ex} v_{nH}$, $1/T_2 = 1/T_D + \sigma_{ex} v_{nH}/2$, n_H is the total density of H atoms which is the same in and outside the maser bulb and

$$I_0 = hV_c / (4\pi\mu_B^2 \eta Q T_D^2)$$

is the threshold flux which is required to turn on the maser, $\eta \approx 0.3$ is the filling factor, $Q=1700$ is the loaded Q of the cavity and $V_c = 12 \text{ cm}^3$ is the cavity volume. Then a flux of at least $I_0 = 6 \times 10^{10} \text{ s}^{-1}$ is required for masing to occur, corresponding to a polarized density difference $\delta_{ac}^0 = 7 \times 10^9 \text{ cm}^{-3}$.

We could measure the total density n_H in the maser bulb by pulsed magnetic resonance with the state selector turned off. We did this with the microwaves on by shifting the magnetic field of the state selector slightly so that the microwaves were no longer at the ESR resonance. This meant that any heating due to the microwaves and also the effects of magnetic forces were unaltered, but the polarization in the maser bulb corresponded to a thermal distribution. The density at threshold was about 10^{11} cm^{-3} suggesting a polarization δ_{ac}^0/n_H of about 15%. At the highest density for which masing was observed, $n_H \approx 2 \times 10^{12} \text{ cm}^{-3}$, the power P was $5 \times 10^{-13} \text{ W}$. This density corresponds to a spin exchange time $\sigma_{ex} v_{nH} \approx T_D$, and we were able to verify this by measuring T_1 as a function of n_H .

Because of cryogenic problems, it was difficult to stabilize the temperature of the state selector which led to oscillations in the amplitude of the maser output. Such oscillations cause frequency fluctuations because of cavity pulling and also through the spin exchange shift. Deliberate mistuning of the cavity can be used to cause these two shifts to cancel. However this version of the apparatus was not equipped for such fine tuning. We have also found the split ring resonator to be somewhat microphonic mainly due to mechanical instabilities of the variable tuning and coupling assemblies so that mechanical vibrations could lead to frequency instability via cavity pulling. We also

note that the use of the small bulb volume to fit the split ring resonator, rather than the larger volume which can be accommodated by a conventional 1420 MHz cavity, results in a reduction in the theoretical stability of the maser by at least a factor of 10.

In spite of these difficulties, we were able to show that the stability of the CHM was at least as good as the stability of a high quality Oscilloquartz B 5400 quartz oscillator on short term loan from the NBS. At $1 \text{ s} \leq \tau \leq 300 \text{ s}$ the Allan variance was 6×10^{-13} equal to the NBS measurements⁷ on the oscillator. Thus the frequency fluctuations of the CHM could not be observed at this level.

One can estimate the stability of our CHM due to a combination of thermal and receiver noise for the parameters of the present experiment

$$\frac{\Delta f}{f} = \left[\frac{kT}{2PQ_L^2 \tau} + \frac{Bk(T + T_N)}{P_A(\omega_0 \tau)^2} \right]^{1/2}$$

where $Q_L = \pi f_{ac} T_2$ is the Q of the maser line, B is the effective noise bandwidth, T_N is the noise temperature of the amplifier and P_A is the power delivered to the amplifier. At a typical power level, $P = 10^{-13} \text{ W}$, P_A was about $0.4 \times 10^{-13} \text{ W}$, and T_N was about 20K. Taking $T_2 \approx T_D = 0.6 \text{ s}$ and $B = 10 \text{ Hz}$, we obtain

$$\begin{aligned} \frac{\Delta f}{f} &= \left[\left(\frac{2.4 \times 10^{-15}}{\sqrt{\tau}} \right)^2 + \left(\frac{3.0 \times 10^{-14}}{\tau} \right)^2 \right]^{1/2} \\ &= 8 \times 10^{-17} \quad \text{for } \tau = 1000 \text{ s.} \end{aligned}$$

The predicted frequency stability of our prototype CHM and of a CHM with a full size cavity, operated at optimum flux, are compared in Fig.3 with the measured frequency stability of two state-of-the-art oscillators, namely the conventional hydrogen maser³ and the superconducting-cavity stabilized oscillator⁸. Also included are the frequency fluctuations we measured by comparing the CHM to the quartz oscillator. This is an upper limit of the Allan variance of our prototype CHM. It is clear

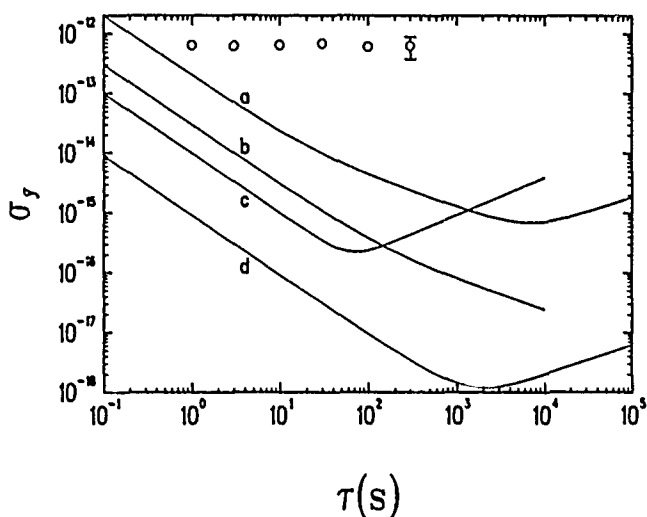


FIG. 3. Plot of relative frequency fluctuations vs averaging time τ : a. measured performance of a conventional hydrogen maser³, b. predicted stability of the prototype CHM with parameters discussed in text, c. measured performance of the superconducting cavity stabilized oscillator⁸, d. predicted stability of a CHM with a full size maser bulb and operated at optimum flux². The circles were obtained by comparing the CHM with the quartz oscillator.

that in order to properly evaluate the performance of the CHM we will need better reference clocks. There are a number of obvious and relatively straightforward improvements that can be made in the CHM. Given these improvements, we are hopeful that the CHM will surpass the performance of conventional masers.

ACKNOWLEDGEMENTS

This work has been supported by a Precision Measurements Grant from the U.S. National Bureau of Standards and by grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. M. Morrow, R. Jochemsen, A.J. Berlinsky, and W.N. Hardy, *Phys. Rev. Lett.* **46**, 195 (1981).
2. A.J. Berlinsky and W.N. Hardy, *Proceedings of the 13th Annual Precise Time and Time Interval Conference*, 547 (1982).
3. R.F.C. Vessot, M.W. Levine, and E.M. Mattison, *Proceedings of the 9th Annual Precise Time and Time Interval Conference*, 549 (1978).
4. W.N. Hardy and L.A. Whitehead, *Rev. Sci. Instr.* **52**, 213 (1981).
5. B.W. Statt, A.J. Berlinsky, and W.N. Hardy, *Phys. Rev. B* **31**, 3169 (1985); B.W. Statt, W.N. Hardy, A.J. Berlinsky, and E. Klein, *J. Low Temp. Phys.* **61**, 471 (1985); and A.J. Berlinsky, W.N. Hardy and B.W. Statt (preprint).
6. M.D. Hürlimann, W.N. Hardy, A.J. Berlinsky and R.W. Cline (submitted to *Phys. Rev. A* (RC)).
7. F. Walls, private communication.
8. S.R. Stein, *Cryogenics*, 363 (July 1980).

SINGLE-STATE SELECTION SYSTEM
FOR HYDROGEN MASERS

Edward M. Mattison*, Robert E.C. Vessot*, and Shen Wei†‡

*Smithsonian Astrophysical Observatory

†Shaanxi Observatory

Academia Sinica

People's Republic of China

SUMMARY

We have designed and operated a single state selection system based on the adiabatic fast passage technique for atomic state inversion. The system uses a variable-pitch DC solenoid and orthogonal 4-wire RF coil to create the state-changing magnetic fields, and a half-wavelength state selection magnet to remove the inverted state atoms from the hydrogen beam. The system's performance was determined by measuring maser power as a function of line Q. The system removes more than 85% of undesired atoms from the beam. We have calculated the potential effects of improvements in the maser on its frequency stability. By using single state selection combined with RF coupling to the cavity and increased storage bulb collimation, stability in the 10^{-17} regime can be achieved for intervals on the order of 6×10^4 seconds.

INTRODUCTION

In the hydrogen maser transitions occur from the ($F=1, m_F=0$) hyperfine ground state of the hydrogen atom to the ($F=0, m_F=0$) state. Only atoms in the ($F=1, m_F=0$) state contribute to maser oscillation, while the presence in the maser storage bulb of atoms in the other three magnetic sublevels of the ground state cause undesired spin exchange relaxation of the radiating atoms and frequency shifts¹ of the output signal. We have constructed and tested a system for removing from the maser's atomic beam all but the ($F=1, m_F=0$) atoms, thus minimizing these undesirable effects.

The state selection system usually employed in hydrogen masers uses a single inhomogeneous state-selection magnet. With this system atoms in both of the upper hyperfine levels, ($F=1, m_F=1$ and $m_F=0$), enter the storage bulb. One can reduce spin exchange relaxation, and reduce the likelihood of frequency shifts due to changes in the hydrogen beam state distribution, by eliminating from the beam the atoms in the ($F=1, m_F=1$) state. In this paper we describe the single-state selection system that we have designed and built to accomplish this goal, and discuss the results obtained with the system. The system utilizes the "adiabatic fast passage" method for changing atomic states, and employs an additional state selection magnet that provides optimum beam optics for this purpose.

SINGLE STATE SELECTION TECHNIQUE

In a single state selection system, shown schematically in Fig. 1, the atoms emerging from the usual first state selection magnet pass through a state-changing region where atoms in the state labelled 1 are changed to state 3, while atoms in state 2 (the upper masing state) are unchanged. The atoms then pass through a second state selection magnet that removes the state 3 atoms, producing the desired beam containing only atoms in state 2.

The beam paths through the state selection magnets are indicated in Fig. 1. The magnets act as velocity filters. The paths of the fastest and slowest atoms that can pass through the system are shown. The range of velocities passed by the single-state system is about the same as in the normal

state selecting system. The beam stops and apertures ensure that atoms in the wrong spin state, and other spurious atoms and molecules, are eliminated from the beam.

The heart of the single-state selection system is the state-changing region. Two general approaches have been suggested for inverting the state-1 atoms into state 3. In the Majorana method² the atoms pass through a region of weak DC magnetic field whose direction is abruptly reversed, interchanging states 1 and 3. The adiabatic fast passage (AFP) technique^{3,4}, which is utilized in the present work, employs orthogonal RF and DC magnetic fields. The DC field varies in strength along the path of the beam, producing in a reference frame rotating with the RF field a total effective field that reverses its direction, causing the precessing atoms to invert their spins.

To produce adiabatic inversion of the $F=1, m_F=1$ state, the magnetic fields must satisfy two criteria. First, the atom's magnetic moment must follow the field direction. Therefore the rate of precession of the atom about the effective field must be greater than the rate at which the field's direction changes with time as the atom moves through it. Second, the effective fields at the ends of the state-changing region must be in approximately opposite directions, so that the atomic states undergo complete inversion. This means that the DC field must vary from weaker to stronger than the RF field (or vice versa) as the atoms traverse the state changing region.

In previous experiments with AFP systems^{3,4} the spatially varying DC magnetic field was created by a permanent ferromagnet and was oriented perpendicular to the direction of beam propagation, while the RF magnetic field was produced by a solenoid parallel to the beam. In the present work the DC magnetic field is produced by a novel technique that allows careful design of the field's spatial variation and easy control of the field's magnitude. The DC field, which is oriented parallel to the beam, is created by a 4-inch (10 cm) long solenoid with a variable-pitch winding. This coil produces a DC magnetic field with a gradient of $1.6 \text{ gauss cm}^{-1} \text{ amp}^{-1}$, and a field at its center of 10 gauss/ampere. Thus a 100 mA DC current produces a central field of 1 gauss and a gradient of 0.16 gauss/cm . The RF field, which is perpendicular to the beam, is produced by an elongated "four-wire" coil placed within the variable-pitch solenoid. The RF coils are 0.64 inches wide and 4.19 inches long, and are spaced 0.32 inches apart. The RF field magnitude is approximately uniform along the length of the state-changing region. The RF field magnitude B_1 is measured by a rectangular pickup coil mounted in the 4-wire RF coil. For a field frequency of 1 MHz, the field measured by the pickup coil is related to the peak-to-peak voltage across it by $B_1(\text{gauss}) = 2.19 V_{p-p}(\text{volts})$. All RF field strength measurements are made at a frequency of 1 MHz.

The solenoids are located within a four-layer cylindrical magnetic shield that isolates the state-changing region from the earth's magnetic field. The components of the single-state selection system are shown in Fig. 2.

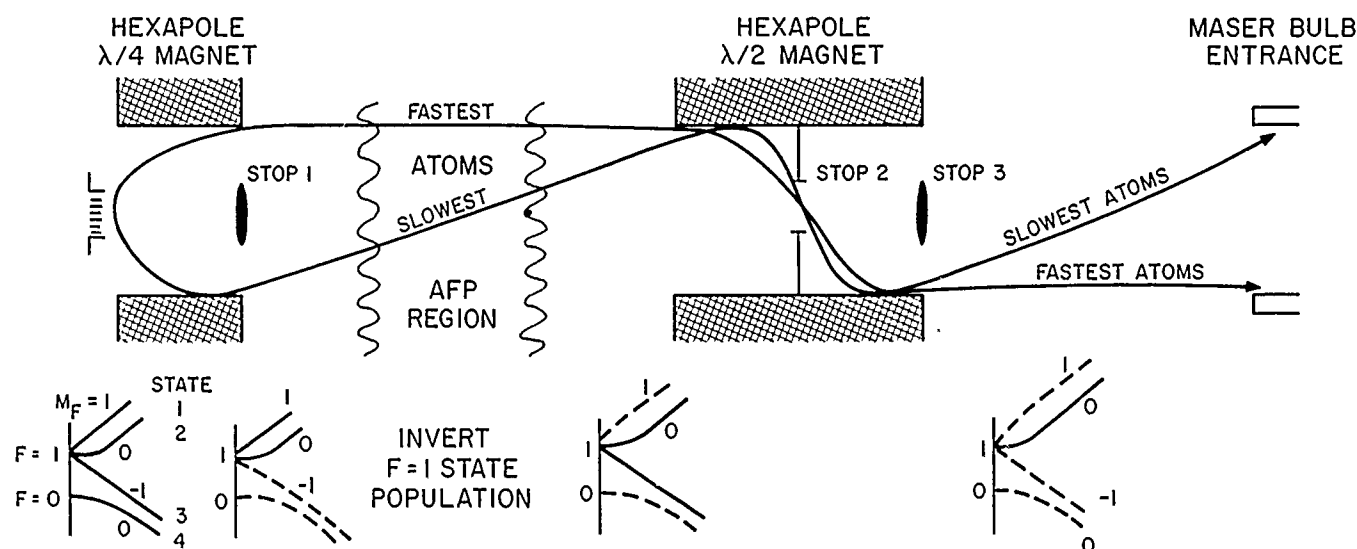


Fig. 1. Schematic diagram of single-state selection system

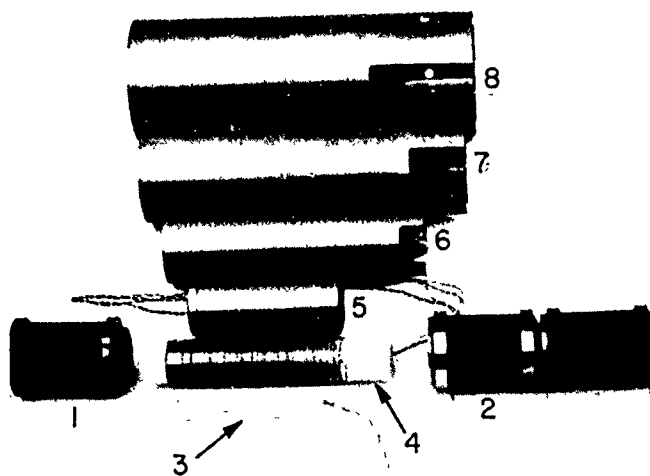


Fig. 2. Components of single-state selection system. 1. First state selection magnet. 2. Final state selection magnet. 3. 4-wire RF coil. 4. Variable pitch DC solenoid. 5-8. Magnetic shields.

The single-state selection system was installed in SAO maser serial number P0. The state-changing system and the second state-selection magnet are suspended from a double-sided vacuum flange that mounts on the maser's normal source flange. A pumping manifold raises the hydrogen dissociator above the state-changing apparatus, and a bellows between the manifold and the dissociator permits alignment of the hydrogen beam with the second magnet.

CHARACTERIZATION OF AFP SYSTEM PERFORMANCE

Maser parameters I/I_{tot} and q

The purpose of a single-state selection apparatus is to ensure that all atoms entering the maser's storage bulb are in state 2 ($F=1$, $m_F=0$). A measure of the performance of such a system, then, is the ratio I/I_{tot} where I is the flux of atoms in state 2 entering the bulb and I_{tot} is the total flux of atoms in all states entering the bulb. Perfect operation of

the system gives $I/I_{\text{tot}}=1$; if atoms in states other than state 2 pass through the system, I/I_{tot} is less than 1. In the absence of state selection the ground state atoms are equally distributed among the four magnetic sublevels, and $I/I_{\text{tot}}=1/4$. (Equivalently, $I_{\text{tot}}/I=4$) The change in the ratio I/I_{tot} is determined indirectly through a measurement of maser line Q as a function of RF power output.

An important parameter describing maser performance is the quantity " q ", defined by ^{5,6}

$$q = \frac{\sigma \bar{v}_r h}{8\pi\mu_0^2} \frac{\gamma_t}{\gamma_d} \frac{V_c}{\eta V_b} \frac{1}{Q_c} \frac{I_{\text{tot}}}{I} \quad (1)$$

Here σ is the hydrogen-hydrogen spin exchange cross section, \bar{v}_r is the average relative velocity of atoms in the storage bulb, $2\pi h$ is Planck's constant, and μ_0 is the Bohr magneton. γ_d is the relaxation rate due to escape from the bulb and recombination on the bulb walls, while γ_t is the total density-independent relaxation rate. If magnetic relaxation is negligible, $\gamma_t/\gamma_d \approx 1.6$. V_c and V_b are the volumes of the maser's resonant cavity and storage bulb, η is the magnetic filling factor^{5,7}, and Q_c is the loaded cavity Q .

All of the quantities on the right side of Eq. (1) other than I_{tot}/I can be calculated or measured, therefore a measurement of q determines the ratio I_{tot}/I . The values used are: $\sigma=2.33 \times 10^{-15}$ cm² (average of σ^+ and σ^- for $T=323^\circ\text{K}$, given by Allison⁸); $\bar{v}_r = 4[kT/\pi m]^{1/2} = 3.69 \times 10^5$ cm/sec for $T=323^\circ\text{K}$; $h = 1.054 \times 10^{-27}$ erg/sec; $\mu_0 = 0.927 \times 10^{-20}$ erg/gauss; $\eta = 2.14$ for a storage bulb radius of 8.839 cm and equivalent cavity length of 2685 cm⁷; and $\gamma_t/\gamma_d = 1$. For maser P0, $V_c = 15184$ cm³, $V_b = 2893$ cm³, and $Q_c = 3.9 \times 10^4$.

q and γ_t can be determined⁶ by measuring the maser's line Q as a function of RF power radiated by the hydrogen atoms into the resonant cavity.

Measurement of the Parameter q

The RF power radiated into the microwave cavity is given by⁵

$$\frac{P}{P_c} = -2q^2 \left(\frac{I}{I_{th}} \right)^2 + (1-3q) \left(\frac{I}{I_{th}} \right) - 1 \quad (2)$$

and the atomic line Q is given by

$$Q_\ell^{-1} = \frac{2\gamma_t}{\omega} \left[1 + q \frac{I}{I_{th}} \right] \quad (3)$$

where $P_c = \frac{\omega^2 V_c}{8\pi\mu_0^2 Q_c \eta}$, $\gamma_t^2 = E\gamma_t^2$ and $I_{th} = \frac{2P_c}{h\omega}$. Eqs. (2) and (3) can be combined to give

$$P = a_2 Q_\ell^{-2} + a_1 Q_\ell^{-1} + a_0 \quad (4)$$

where $a_2 = -\frac{E\omega^2}{2}$ (5)

$$a_1 = \frac{\omega E \gamma_t}{2} \left(1 + \frac{1}{q} \right) \quad (6)$$

$$a_0 = -\frac{E\gamma_t^2}{q} \quad (7)$$

If the values of a_1 and a_0 are known, Eqs. (6) and (7) can be solved for q and γ_t :

$$q = -\frac{(2+\varphi)}{2} \left[1 - \sqrt{1 - \frac{4}{(2+\varphi)^2}} \right] \quad (8)$$

where $\varphi \equiv \left(\frac{2a_1}{\omega} \right)^2 \frac{1}{a_0 E}$ (9)

and $\gamma_t = \frac{2a_1}{\omega E} \frac{q}{1+q}$ (10)

To determine q and γ_t , measurements of line Q are made for several values of beam flux, and therefore of power P . The data can be treated in either of two ways. If the power is high enough for the data to show significant curvature, a quadratic form may be fit to the data and the coefficients a_2 , a_1 , and a_0 can be determined. a_2 provides an independent estimate of the value E , and a_1 and a_0 yield q and γ_t . If the maser power is low, the curvature of the data is often insufficient to allow a_2 to be determined accurately. In that case the value of a_2 is calculated a priori from the quantities making up E , and a new function $F(Q_\ell) \equiv P - a_2 Q_\ell^{-2}$ is calculated. A straight line of the form $(a_1 Q_\ell^{-1} + a_0)$ is fit to $F(Q_\ell)$ to determine the coefficients a_1 and a_0 , from which q and γ_t are calculated using Eqs. 8-10.

SINGLE-STATE SELECTION EFFICIENCY

The effectiveness of the single state selection system is determined by comparing the values of q measured with the AFP system on and off. In order to determine whether the presence of the state selection apparatus (without the RF field operating) mixes the atomic states and reduces q , we measured q with the maser in its standard configuration, without the single state selection system installed. Fig. 3 shows the measured maser power as a function of inverse line Q . The quadratic nature of the relationship is evident. The solid curve is the parabola that represents the best least-square fit to the data

points. The value of a_2 resulting from the linear regression is $a_2 = -3.476 \times 10^{13}$ erg/sec, which is within 4% of the calculated value of -3.35×10^{13} erg/sec. This confirms the values of Q_c , η , and V_c used in the calculations of a_2 . The values of q and γ_t determined with the single state selector out of the maser are $q = 0.080 \pm 0.003$, and $\gamma_t = (1.0 \pm 0.024) \text{ sec}^{-1}$.

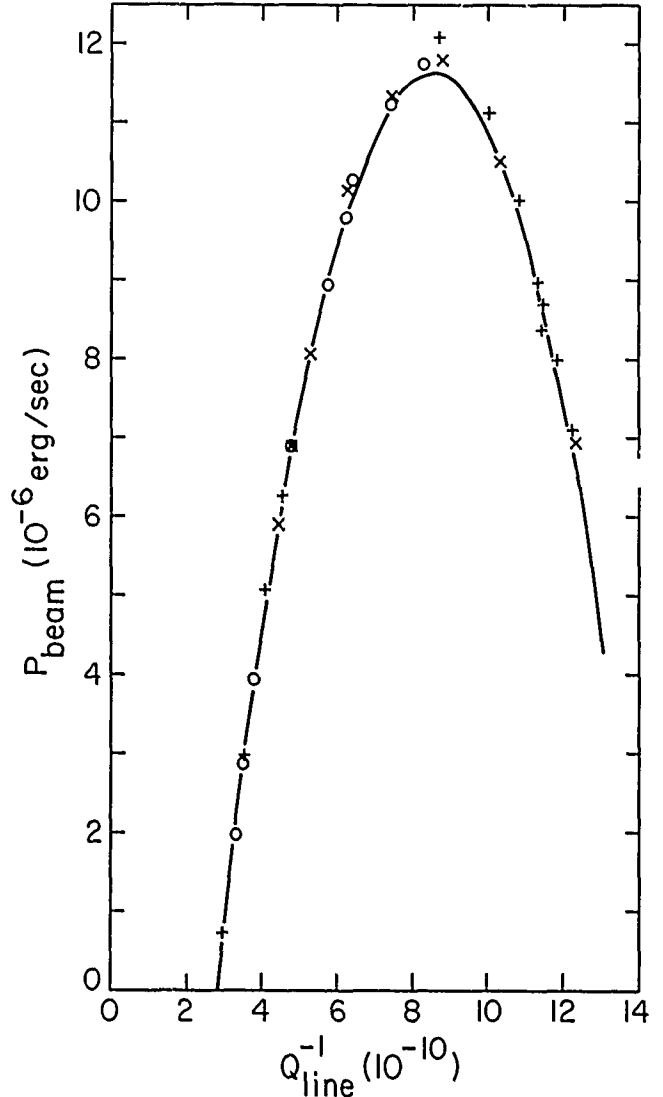


Fig. 3. Maser power vs. inverse line Q with single state selection system not installed in maser

With the single state selection system installed in the maser, q was measured for several values of the AFP magnetic fields, with the RF field both on and off. Fig. 4 is a typical plot of maser power against inverse line Q . Because the available beam flux is limited in the experimental single-state selection apparatus (due to the length and limited pumping speed of the intentionally flexible design), a relatively small and approximately linear portion of the quadratic curve is observed. For this reason the coefficient a_2 is not well determined by the data, and we use the alternate method of calculating q and γ_t . Therefore Fig. 4 shows a plot of $P_{beam} - a_2 Q_{line}^{-2}$, with $a_2 = -3.35 \times 10^{13}$ erg/sec. a_1 and a_0 are determined from straight-line fits to the data, and q and γ_t are calculated from a_1 , a_0 , and the assumed value of a_2 .

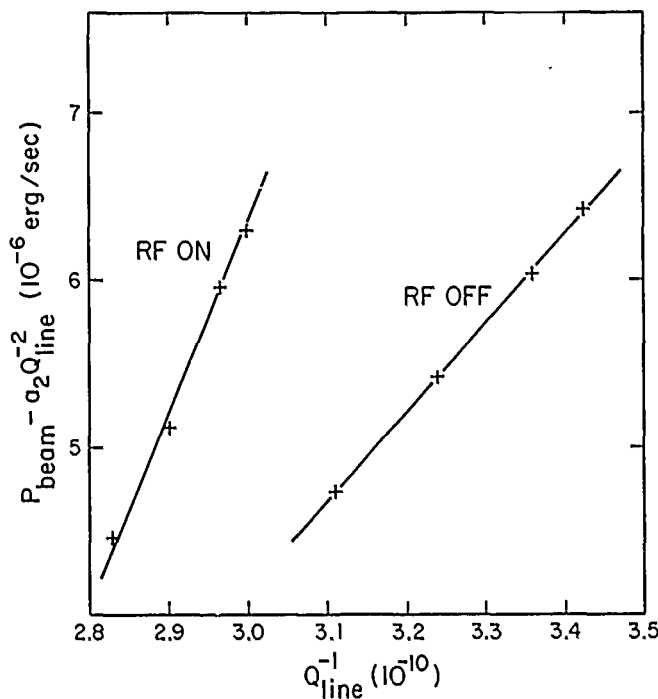


Fig. 4. Linearized maser power vs. inverse line Q.

Table 1 summarizes the experimental conditions and results of q measurements. From Table 1 it can be seen that there is close agreement between the values of q with the single-state selection apparatus not installed (run 1, Fig. 3) and with the system installed but with the AEP RF magnetic field turned off (run 2, Fig. 4).

Table 1. Results of q Measurements

BDC (C)	BRF (C)	f _{RF} (MHz)	q _{on}	q _{off}	Γ	γ_t (sec ⁻¹)	run
Apparatus not installed				.080	--	1.10	1
1.0	0.55	1.4	.039	.081	2.08	1.10	2
1.0	0.57	2.65	.041	.070	1.71	1.15	3
1.0	0.41	3.22	.040	.070	1.75	1.16	3
1.0	0.41	1.20	.042	.070	1.67	1.15	3
1.0	0.45	1.3	.041	.071	1.74	1.15	4
0.5	0.45	1.22	.043	.072	1.67	1.13	5

This agreement verifies that the presence of the single state selection system does not significantly perturb the state distribution in the atomic beam when the system is turned off. Thus the ratio $\Gamma \equiv (I_{\text{tot}}/I)_{\text{off}} / (I_{\text{tot}}/I)_{\text{on}}$ measures the fractional improvement in the beam purity compared with the maser using the usual state-selection magnet. The value of Γ for the case of run 2, which equals 2 to within experimental error, indicates that essentially all of the level 1 atoms have been eliminated from the beam. The data of runs 3-5 give values of $q(\text{AEP off}) \approx 0.070$ and $\Gamma \approx 1.7$. This value of Γ implies that approximately 85% of the level 1 atoms have been eliminated from the beam. The discrepancy between the data of runs 3-5 and those of runs 1 and 2 may be due to the fact that different receivers, with slightly different calibration procedures, were used for the two sets of data. This could have resulted in an error of up to 0.5 dB in the absolute power measurement for the data of runs 3-5. By comparison, the receiver used for the runs 1 and 2 was calibrated by direct comparison with a Marconi model 6960 RF power meter, resulting in a confidence of better than 0.2 dBm in its absolute power measurement. This

accuracy is reflecting in the agreement between the calculated coefficient a_2 and that measured from the Fig. 3 (run 1) data.

Within the calibration errors in the experiment, the data show that the single-state selection employing adiabatic fast passage state-changing eliminates at least 85% of the undesired level 1 atoms from the beam. A redesigned apparatus with greater pumping speed and higher beam flux would be highly desirable for operating masers. Such a system would allow use of a storage bulb with a longer storage time constant than is employed at present. (If the bulb storage time is increased without decreasing the flux of level-1 atoms in the beam, the increase in line Q would soon be limited by spin-exchange relaxation.) The combination of single-state selection and longer bulb time constant would allow a substantial increase in the maser's line Q, and thus in its long-term frequency stability.

IMPLICATIONS for MASER FREQUENCY STABILITY

The major benefit to be gained from single-state selection is an increase in maser frequency stability. We estimate the stability improvement that can be gained by use of single-state selection, modifications to the storage bulb, and changes in the maser's operating conditions. The maser's frequency stability, as expressed by the Allan variance, can be written as

$$\sigma^2(\tau) = \sigma_S^2(\tau) + \sigma_M^2(\tau) + \sigma_L^2(\tau) \quad (11)$$

$$\text{with } \sigma_S^2 = \frac{FkTB}{\omega^2 P_{\text{ext}} \tau^2} = \left[\frac{FkTB}{\omega^2 \tau^2} \right] \left(\frac{1+\beta}{\beta} \right) \frac{1}{P} \quad (12)$$

$$\sigma_M^2 = \left[\frac{kT}{2\tau} \right] \frac{1}{PQ_L^2} \quad (13)$$

$$\sigma_L^2 = 9 \times 10^{-36} \tau \quad (14)$$

Here F is the noise figure of the maser's receiver system, k is Boltzmann's constant, T is the absolute temperature of the maser's storage bulb, B is the equivalent noise bandwidth of the receiver, P_{ext} is the power delivered to the receiver, β is the coupling factor to the cavity, P is the total power radiated by the hydrogen atoms in the storage bulb, and τ is the averaging interval during which σ is measured.

σ_S , the short-term Allan variance, represents the frequency fluctuations due to additive phase noise at the receiver input. It is proportional to τ^{-1} and dominates the stability for averaging intervals less than roughly 50 seconds. σ_M , the medium-term stability, results from thermal noise within the linewidth of the atomic oscillator. It is proportional to $\tau^{-1/2}$, and is important at intervals between roughly 50 seconds and 3000 seconds. σ_L is an empirical representation of the frequency stability over time intervals longer than approximately 4000 seconds⁹. Long-term frequency variations are most likely due to a variety of systematic processes, rather than to thermal noise. The long-term stability of a particular maser depends upon its construction and environment. The form of σ_L given in Eq. (14) has been observed for pairs of VLG-11 masers. We assume that most of the systematic effect is due to pulling of the atomic resonance line by the cavity resonance, and thus assume that the long-term Allan variance is proportional to Q_C/Q_L , the ratio of cavity Q to atomic line Q.

Numerical values for the coefficients in Eqs. 12 and 13 can be calculated assuming the following quantities used in practice: $F=4$ (=6dB), $k=1.38 \times 10^{-16}$ erg/°K, $T=322^\circ\text{K}$, and $B=6\text{Hz}$.

$$\sigma_S = 1.16 \times 10^{-16} \left[\frac{1+\beta}{\beta} \right]^{\frac{1}{2}} P^{-\frac{1}{2}} \tau^{-1} \quad (15)$$

$$\sigma_M = \frac{1.49 \times 10^{-7}}{Q_L B^{\frac{1}{2}}} \tau^{-\frac{1}{2}} \quad (16)$$

The coefficient in Eq. 41 corresponds to $Q_L \sim 2 \times 10^9$ and $Q_C \sim 39000$, allowing us to write for illustrative purposes,

$$\sigma_L = 1.5 \times 10^{-13} \frac{Q_C}{Q_L} \tau^{\frac{1}{2}} \quad (17)$$

P and Q_L can be expressed in terms of measurable and controllable quantities. The line Q is given by

$$Q_L = \frac{\omega}{2(\gamma_t + \gamma_{se})} \quad (18)$$

where γ_{se} is the portion of γ_2 due to spin exchange.

Also
$$\gamma_{se} = \frac{\bar{v} \sigma_{se}}{v_b} \left(\frac{I_{tot}}{I} \right) \frac{I}{\gamma_t} \quad (19)$$

Therefore a reduction in I_{tot}/I reduces γ_{se} , while a reduction in γ_t (due, for example, to increased collimation of the storage bulb) increases γ_{se} for a constant flux I of ($F=1$, $m_F=0$) atoms. Combining Eq. 18 with Eq. 2 for the beam power P , we obtain

$$P = E \left[-2(\gamma_t + \gamma_{se})^2 + \left(1 + \frac{1}{q} \right) \gamma_t (\gamma_t + \gamma_{se}) - \frac{\gamma_t^2}{q} \right] \quad (20)$$

where $E=0.0328 Q_C^{-1}$ erg-sec.

Using Eqs. 15-20 we can calculate the frequency stability for any maser operating condition, and thus examine the effect of adjusting various maser parameters. The results of several scenarios are shown in Fig. 5.

(1.) We first assume that the maser is in its normal condition, with AFP turned off and the flux adjusted for a line Q of 2×10^9 , a typical value. From the measured value $\gamma_t = 1.1 \text{ sec}^{-1}$ and the calculated storage bulb escape rate $\gamma_b = 0.87 \text{ sec}^{-1}$, we deduce that the rate of recombination on the bulb wall is $\gamma_r = \gamma_t - \gamma_b \approx 0.2 \text{ sec}^{-1}$. The calculated frequency stability is consistent with values measured in practice.

(2.) In this case the maser parameters are as in (1), but the AFP is turned on. Single-state selection reduces I_{tot}/I by a factor of 2, halving q and γ_{se} . The effect of single-state selection is to increase both the line Q and the beam power, increasing the overall frequency stability. The short-term Allan variance σ_S is decreased by a factor of 1.2, the medium term variance σ_M by a factor of 1.6, and the long-term variance σ_L by a factor of 1.4.

(3.) In order to increase the maser's power output while still obtaining the high line Q resulting from a long storage time, we can greatly increase the RF coupling to the resonant cavity. The increased coupling directs more power to the receiver, thus improving the short-term frequency stability. It also decreases the loaded cavity Q and therefore the cavity pulling, thus proportionately improving the long-term stability in our model. We assume here that γ_t and γ_{se} are each reduced to 0.4 sec^{-1} , as in case 2, and

that the cavity Q is reduced by a factor of 2, from 39,000 to approximately 20,000, by increasing the coupling factor β from 0.2 to 1.4. The combination of single-state selection and increased cavity coupling results in $q = 0.08$, its original value in case 1. The decreased relaxation rates give a line Q that is 2.8 times its value in case 1, a beam power 25% lower than in case 1, and receiver power 2.7 times greater than in case 1. In addition, the ratio of line Q to cavity Q is increased by a factor of 5.6 over case 1. The result is substantial improvement in the short, medium, and long-term frequency stabilities.

(4.) To examine the frequency stability improvements to be gained from further improvements in the wall coating, we consider the effects of reducing the wall recombination rate by a factor of three. That this reduction might be possible is indicated by low temperature experiments¹⁰ showing that coatings of frozen-in-place CF_4 have wall shifts one-fourth that of Teflon. This lower wall shift implies a smoother surface, which would lead to lower collision rates and thus lower relaxation rates.

We assume here that the storage wall recombination rate is reduced from 0.2 sec^{-1} to 0.07 sec^{-1} , that the bulb escape rate is made equal to 0.07 sec^{-1} , and that the beam flux is reduced to make $\gamma_{se} = \gamma_t = \gamma_b + \gamma_r = 0.14 \text{ sec}^{-1}$. Also, we assume that the cavity coupling is increased as in case 3.

The decreased relaxation rates greatly increase the line Q , but decreases the beam power. The result is a decrease in short-term stability over case 3 by a factor of 5, a decrease in medium-term stability by a factor of 1.7 (but still an improvement over case 1), and an increase in long-term stability by a factor of 2.9 over case 3 and by a factor of 16 over case 1.

Assuming that long-term maser frequency variations are due primarily to cavity frequency pulling effects, the use of single-state selection and storage surface improvements can lead to long-term frequency stabilities in the 10^{-17} regime.

It can be seen that substantial improvements in maser frequency stability can be realized from single-state beam selection, storage bulb collimator adjustment, increased cavity coupling, and improved storage bulb surfaces. By using all of these techniques, long-term frequency stability on the order of 9×10^{-17} can be achieved over intervals on the order of 6×10^4 seconds.

ACKNOWLEDGEMENTS

We thank the Jet Propulsion Laboratory and the U.S. Naval Research Laboratory for their support of this work, which was performed under JPL contract 956950 and NRL contract N00014-79-C-0718. We are pleased to acknowledge the invaluable help of Mr. Donald Graveline of the Smithsonian Astrophysical Observatory, who designed and assembled the vacuum housing for the experimental apparatus.

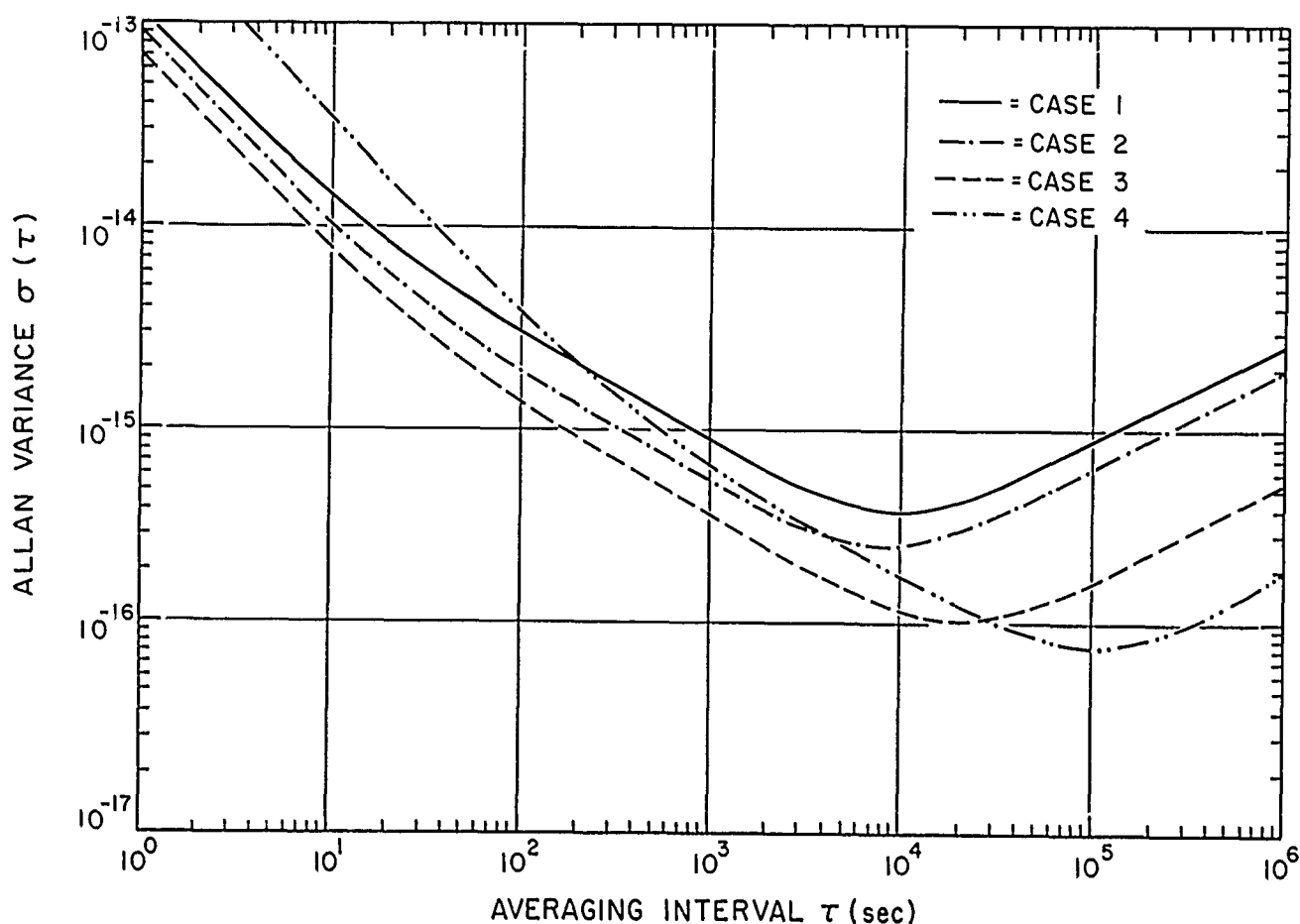


Fig. 5. Graph of Predicted Frequency Stability

REFERENCES

1. Crampton, S. B., and H. T. M. Wan, "Duration of hydrogen-atom spin-exchange collision", *Phys. Rev. A* **12**, 1305 (1975).
2. Urabe, Shinji, Koji Nakagiri, Yasusada Ohta, Masaki Kobayashi, and Yoshikazu Saburi, "Majorana effect on atomic frequency standards", *IEEE Trans. on Instr. and Meas.* **IM-29**, 304 (1980).
3. Lacey, Richard F., and Robert R. C. Vessot, "Improved state selection for hydrogen masers", in *Proc. 23rd Freq. Control Symp.*, p. 279 (1969).
4. Audoin, Claude, Michel Desaintfuscien, Pierre Petit, and Jean-Pierre Schermann, "Design of a double focalization in a hydrogen maser", *IEEE Trans. on Instr. and Meas.* **IM-17**, 351 (1968).
5. Kleppner, D., H.C. Berg, S.B. Crampton, N.F. Ramsey, R.F.C. Vessot, H.E. Peters, and J. Vanier, "Hydrogen maser principles and techniques," *Phys. Rev.* **138**, A972 (1965).
6. Mattison, Edward M., Wei Shen, and Robert F.C. Vessot, "Measurement and interpretation of hydrogen maser quality parameters", in *Proc. 39th Ann. Freq. Control Symp.*, p. 72 (1985).
7. Kleppner, Daniel, H. Mark Goldenberg, and Norman F. Ramsey, "Theory of the hydrogen maser." *Phys. Rev.* **126**, 603 (1962).
8. Allison, A.C., "Spin-change frequency shifts in H-H collisions," *Phys. Rev. A* **5**, 2695 (1972).
9. Vessot, R.F.C., M.W. Levine and E.M. Mattison, "Comparison of theoretical and observed hydrogen maser stability limitation due to thermal noise and the prospect for improvements by low-temperature operation," in *Proc. 9th Ann. Precise Time and Time Interval Applications and Planning Meeting*, p. 549 (1977).
10. Vessot, R.F.C., E.M. Mattison, E.A. Imbier, and Z.C. Zhai, "Atomic hydrogen maser measurements with wall surfaces of carbon tetrafluoride," in *Proc. 16th Ann. Precise Time and Time Interval Applications and Planning Meeting*, p. 357 (1984).

PROGRESS TOWARD AN OPTICALLY PUMPED CESIUM BEAM FREQUENCY STANDARD*

R.E. DRULLINGER, JON SHIRLEY, D.J. GLAZE, L.W. HOLLBERG
National Bureau of Standards
Boulder Colorado 80303 and

A. DeMARCHI
Istituto Elettrotecnico Nazionale G. Ferraris
Torino Italy

Abstract

The National Bureau of Standards is planning to build a cesium-beam, primary frequency standard based on the application of optical pumping for state selection and atom detection. The goal is an accuracy of 10^{14} .

Theoretical studies have been able to identify only Rabi pulling as a mechanism for Majorana-transition-induced frequency shifts. Together with considerations of magnetic field uniformity, this has led us to adopt a longitudinal C-field. In turn, this has required a hybrid magnetic shield design with an active component to cancel ambient fields. Elimination of state-selecting magnets together with polarization control of the optical pumping should eliminate effects of Majorana transitions.

Optical pumping should also permit simultaneous operation of counter-propagating beams with closer trajectory retrace than is possible with magnetic state selection. Real-time measurements of end-to-end cavity phase shift and even servo control are anticipated. Requirements on distributed cavity phase shift have led to consideration of a "race track" shaped cavity termination in place of the more conventional shorted waveguide.

Noise measurements have shown that simple monolithic diode lasers produce too much FM noise to allow one to reach the shot noise limit in atom detection. Techniques for control of diode noise and linewidth are being tried and compared.

Introduction

The possibility for realizing cesium atomic beam frequency standards in which the state preparation and subsequent detection are accomplished by means of laser driven optical pumping and resonance fluorescence has been studied for several years both theoretically and in the laboratory.¹ This technology offers so much potential for improving the evaluation and/or control of several of the more serious accuracy-limiting systematic errors found in conventional beam standards that the National Bureau of Standards has begun a project to build a large prototype standard in which we can investigate the realizable limits of the technology. After the prototype is used to study some basic effects like frequency shifts due to fluorescent light and blackbody radiation, Majorana effects and distributed-cavity phase shift in new cavity geometries it is expected that this device can become the next NBS primary standard with only minor modifications. For this reason the design specifications are those of a standard in which the accuracy should be about 10^{14} .

The major building blocks of such a standard are (aside from electronics and laser systems): microwave cavity, magnetic shields, fluorescence collection optics and the vacuum envelope. However, their designs are highly interdependent. The design logic has to be based on the interplay of the major systematic error causing effects found in atomic beam frequency standards.

Magnetic Field and Rabi Pulling

The requirements on C-field homogeneity and stability needed to control the quadratic Zeeman shift are severe unless small field values can be used. On the other hand, shifts caused by the overlapping wings of other Zeeman transitions (Rabi pulling) are reduced by working at higher fields. The tradeoff can be facilitated by adopting a longitudinal C-field. This geometry will result in better field homogeneity. Also, in this geometry, most microwave cavities contain a field pattern which varies as a half-sine-wave along the atomic beam rather than the more common rectangular pulse. Theoretical studies have shown that the associated Rabi line shape is smoother and much smaller in the wings than in the case of the rectangular pulse², thereby reducing Rabi pulling. This will allow one to work at the lower C-field values required by stability considerations without excessive Rabi pulling.

However, the longitudinal C-field is not without drawbacks. Long cylindrical shields have a smaller shielding factor in the axial direction³; the direction in which the shielding now becomes most critical. This has forced us to consider a hybrid shield package in which an actively servoed coil outside the passive shields will be used to buck the ambient longitudinal field to near zero.

It may also be possible to actively servo the mean C-field value by monitoring the frequency of the first field sensitive transition. The microwave power can be modulated to put a small amount of power into sidebands at the separation of these transitions. With an additional audio modulation, different from that of the main clock servo and a separate phase sensitive detection channel in parallel with the main servo, one could monitor and control the effective C-field. Because this need not be a tight or fast servo loop, the time constant could be quite long and the amount of signal (noise) insignificant with respect to the signal from the main clock transition.

Majorana Effects

The theoretical effort has focused on Majorana transitions and Rabi pulling. Rabi pulling, as

*Contribution of NBS, not subject to copyright.

pointed out above, can result when the wings of the Rabi pedestals of other Zeeman components in the hyperfine spectrum overlap the clock transition. But a net shift occurs only if the distribution of population among the magnetic sublevels is asymmetric, since the spectrum is otherwise symmetric at low magnetic fields. When magnetic state selection is used, the symmetry is broken in the high magnetic fields and atoms with different m values follow slightly different trajectories. Hence, asymmetric populations and the associated Rabi pulling are a common source of error in conventional cesium standards. Majorana transitions can compound the problem by transferring atoms from the more asymmetrically populated high m sublevels to the low m sublevels closest to the clock transition.

Optical state selection by circularly polarized light propagating in the direction of the magnetic field can produce a strong asymmetry in the populations of the Zeeman sublevels. However, either the use of linearly polarized light, or light propagating perpendicular to the magnetic field should produce symmetric populations. Doing both, we expect to achieve highly symmetric populations and a consequent vanishing of Rabi pulling. Thus, optical pumping permits the deliberate introduction of Rabi pulling or its suppression.

Majorana transitions are caused by directional changes in the magnetic field as seen by atoms passing through the apparatus. Elimination of the state-selecting magnets and extension of the C-field and magnetic shields over the optical pumping region should keep the field sufficiently uniform that Majorana transitions will not occur. Even if they do occur, they will not lead to Rabi pulling shifts since the optical pumping produces a symmetric population distribution. Majorana-transition probabilities have the same symmetry properties as rotation operators and can not generate asymmetric populations from initially symmetric ones.

Phase Shifts and Cavity Design

Phase shifts in the microwave cavity can be divided into two classes. One is the variation in phase from one end of the cavity to the other caused by electrically asymmetric lengths of the two arms of the Ramsey cavity and is usually evaluated by beam reversal. The second is the variation in phase across the window in the cavity through which the atomic beam passes. This distributed-cavity phase-shift is relatively small, being caused by propagation losses in the waveguide.⁴ Even though it is not explicitly evaluated in present standards, its existence places limits on one's ability to accurately evaluate the larger end-to-end phase shift.

The end-to-end phase error seen by an atom traversing an imperfect cavity can be evaluated by measuring the frequency difference when atoms are made to traverse the same cavity in exactly the opposite direction. The word same is underlined to emphasize that nothing must happen to the cavity during the reversal that may result in changes in its phase shifts; hence, the beam direction is usually reversed rather than the cavity itself. This evaluation process is further complicated and limited by mechanisms which spatially disperse the atomic beam, generating inhomogeneities in flux density and

velocity across the beam which then vary along the beam. These inhomogeneities create an undesirable sensitivity to distributed-cavity phase shift in the beam reversal process by limiting the precision of the beam retrace. This causes a differential sensing of the distributed phase shifts for the two directions.

Magnetic state selection is a velocity dependent process resulting in a fan shaped beam (assuming dipole optics) with higher velocities being deflected less and slower velocities deflected more. State selection by optical pumping, on the other hand, does not result in significant spatial beam dispersion. This alone should result in a substantial improvement in our ability to evaluate end-to-end phase shift.

An optically pumped laboratory standard offers the additional possibility of monitoring end-to-end cavity phase shift in real time and perhaps servo controlling it.⁵ Figure 1 shows a schematic of a beam tube in which counter propagating beams operate simultaneously through the same cavity. The beam flux is sufficiently low that the two beams do not collisionally interact. The optical state selection/detection process could work as follows.

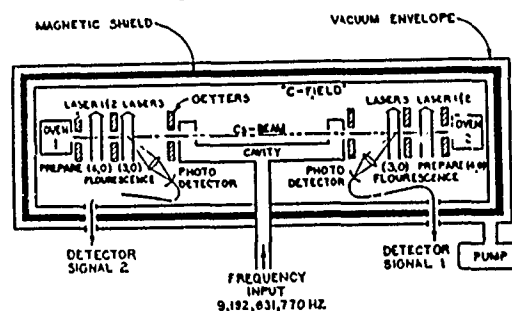


Figure 1. Schematic of a beam tube in which counter propagating beams operate simultaneously.

The beam from the oven on the left side of the figure first passes through a region of optical pumping in which one hyperfine state is essentially depleted of all population. On passing through the Ramsey cavity some population is excited back into that state and subsequently detected by resonance fluorescence in the detection zone on the right end of the cavity. A second oven and optical pumping region on the right with their corresponding detection zone on the left make the machine totally symmetric. The optical detection process is resonant only with atoms in the terminal state of the clock transition and therefore, to the extent that the optical pumping is complete, is virtually blind to atoms that have not yet undergone the clock transition. Furthermore, the subsequent signal processing electronics are sensitive only to the modulation imposed on the signal by the microwaves. The result is that, to a very high degree, each detector is sensitive only to the beam coming from its corresponding oven and not to the simultaneous presence of a counter propagating beam. In this way the frequency error caused by the net cavity phase shift can be monitored in real time. The simple expedient of differentially heating one arm of the Ramsey cavity can then be used to cause differential expansion and drive the error to zero.

Beam reversal, however, treats only the as-perceived end-to-end phase shift without regard to potential sensitivity to residual distributed phase

shift which may creep into the system through non-exact beam retrace. The distributed phase shift caused by propagation losses in conventional shorted waveguide cavities has been analyzed^{4,6} and shows that for an error no larger than 3×10^{-15} (the level we would need in a standard of 10^{-14} overall accuracy), the center of gravity of perfectly homogeneous beams must retrace to better than 100 μm . This is an uncomfortable limit and it has caused us to look for cavity designs in which the effect is smaller.

Ideally, the microwave field at the point of atomic beam passage should be a perfect standing wave. Most waveguide cavities used in the past generated the standing wave by reflecting a traveling wave back upon itself at a short. Since a short is not a perfect conductor, the reflected wave is not quite equal to the incident wave. It is this inequality in the counter-propagating traveling waves that gives rise to the distributed phase shift. We plan to use a racetrack shaped cavity (Fig. 2) whose field more closely approaches the ideal

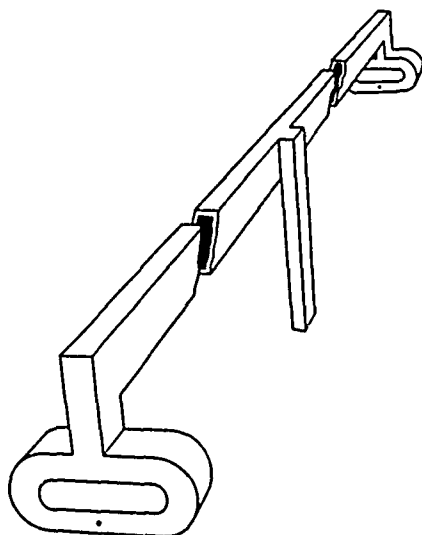


Figure 2. New cavity, proposed to reduce the effects of distributed-cavity phase shift.

standing wave, and hence, produces the desired lower sensitivity to beam retrace errors. An analysis of the distributed-cavity phase shift to be expected in this cavity (including effects attributable to imperfect fabrication) is provided in an accompanying paper.⁷ We have made several test pieces both by machining the cavity into a solid copper block and by assembling standard microwave pieces. Measurements of cavity Q, resonance frequencies and imperfection induced coupling to undesirable modes have confirmed the model.

A Horizontal Beam Tube and Gravitationally Induced Beam Dispersion

In horizontal beam tubes, the effect of gravity is to disperse the atomic beam with slow atoms falling more than fast ones over the length of the machine. In long standards the resulting spread can be several millimeters. To retain the advantages of

the homogeneous beam produced by optical pumping, our new standard must be either relatively short or vertical.

A large vertical machine potentially presents problems with variations in cavity phase shift and shielding produced by temperature changes and gravitational strain effects. It could also suffer from temperature gradients. Such a machine could not easily be made large enough to offer a significant advantage in line Q. Hence, we are designing a horizontal beam tube with a Ramsey cavity of about 1.5 m, significantly shorter than the 3.75 m of NBS-6.

The residual sensitivity to distributed-cavity phase shift caused by the gravitationally dispersed beam can be handled in several ways. Distributed-cavity phase shift by its very nature is oriented along the Poynting vector in the cavity. By orienting the cavity so this is orthogonal to the beam dispersion one gains some immunity from this effect. One might also deflect the atom beam with photon pressure in a way that partially cancels the gravitational dispersion. Using a cycling transition, all velocities can be caused to refocus at some designated point downstream. For a flight path of 2 m, average velocity atoms will require 15 photon scatterings. Certainly one would have no trouble scattering enough photons to accomplish the desired deflection. On the contrary, this number is so small that one must be concerned about the unintentional deflection and/or beam blow up possible in some of the state selection schemes that have been proposed.

Optics, Lasers and Related Topics

The potential systematic error caused by fluorescent light produced in the optical pumping process has already been treated theoretically.⁸ Within the constraints on atomic beam length presented above, the laser beams can be arranged far enough from the Ramsey cavity to keep light shifts acceptably small.

The fluorescence collection optics are shown in figure 3. A pair of spherical mirrors will collect and image the fluorescence onto the end of a light pipe which will relay the light out through the magnetic shields to a silicon photodiode detector

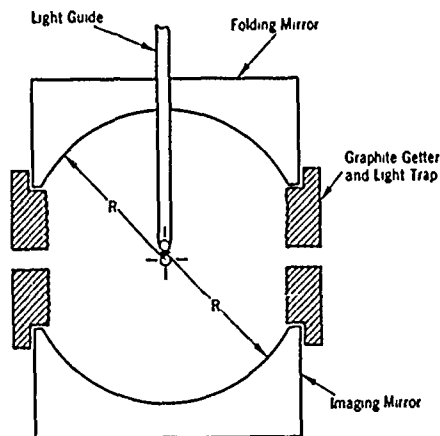


Figure 3. Fluorescence collection optics.

outside the vacuum chamber. Ray trace programs have been used to model the coupling of light into the end of the light pipe. They show that for small object to image distance ratios, the spherical aberrations are tolerable. Also, light scattered from the end of the light pipe by the Gaussian wings of the laser beam has been considered and appears manageable within the constraint just given on object to image distance. Depending on the light pipe technology used (glass clad or mirrored tube), up to 60% of the fluorescence can be collected. The use of the graphite skirt will reduce both the scattered light and background cesium within the detection area.

Our present work on diode lasers involves trying to achieve the fundamental detection limits imposed on the resonance fluorescence method by the atomic beam shot noise. With simple commercially available diode lasers one finds excess noise. At least one component of this noise on the cesium fluorescence signal is the result of diode laser frequency fluctuations which sweep the laser frequency across the narrow (~ 5 MHz) atomic linewidth.

We see improved signal to noise in the detected fluorescence when the diode laser frequency is actively locked to the cesium atoms or to an external reference cavity. In locking the laser to the cesium atomic fluorescence two methods have been used; 1) the laser is locked directly to the cesium atomic beam fluorescence or 2) the laser is locked to the resonance via a Doppler free saturated absorption signal in a separate cesium cell. All three of these techniques significantly reduce the excess noise on the fluorescence signal, but as of yet the noise is not reduced to the level expected for shot noise limited detection.

The locking systems employed until now reduce the laser frequency jitter at low Fourier frequencies (below ~ 20 kHz) and show a reduction in the amplitude noise on the fluorescence signal. It should be noted that with this low unity gain frequency the fast laser linewidth is not reduced. The present diode laser systems with weak optical feedback have a laser phase stability which is limited to no more than 20 ns. The rapid laser phase fluctuation requires a very fast servo-control system to appreciably reduce the diode linewidth.

Other sources of noise are being studied as are improved implementation of the present frequency control systems.

Design Summary

We are building a prototype optically pumped, cesium-beam standard in which we can study systematic effects not amenable to study in conventional standards. It is our expectation that, after these studies, the device can be made to operate as a new primary standard with an accuracy of 10^{14} or almost an order of magnitude better than our existing standard. At this level of accuracy our time scale will not support the presently practiced policy of evaluating the standard once per year. Rather the new standard will have to be evaluated monthly. To avoid this position, we are planning to make the standard operate at least quasi-continuously as a clock and, to the extent possible, be self evaluating. The design has followed from an analysis of all identified systematic errors.

To a large extent the design need not be committed to any particular optical pumping scheme at this time. Our plan is for a horizontal beam tube with a Ramsey cavity of about 1.5 m. It follows from this value and the mean thermal velocity, that a line Q of about 10^8 is to be expected. To realize the stated accuracy goal will, therefore, place a heavy burden on the line centering servo. Our preliminary investigation (not discussed here) indicates the feasibility of the task.

A new racetrack shaped microwave cavity has been designed and modeled which will reduce distributed-cavity phase shift and requirements on beam retrace precision. Plans call for simultaneous operation of counter-propagating atomic beams and servo control of the end-to-end phase shift. Regions of optical pumping and detecting will be placed symmetrically on each end of the Ramsey cavity separated by 40 and 20 cm respectively from the cavity. With the expected beam flux and pumping schemes, these separations are adequate to control light shifts.

For reasons of magnetic field control as well as Rabi line pulling, a longitudinal C-field has been selected. This in turn has required the use of a hybrid magnetic shield design in which both active and passive shielding will be used.

Work continues on diode laser characterization and control. The goal is to find a simple way to use the diode lasers and still reach the fundamental limits on clock performance. Failing this, we can resort to the much more complicated but certain techniques of external cavity lasers.

References

1. R.E. Drullinger, Proc IEEE 74, 140 (1986) and references cited therein.
2. Jon. H. Shirley, Phys Rev 160, 95 (1967)
A. DeMarchi, G. D. Rovers, and A. Premoli, Metrologia 20, 37 (1984).
3. S. M. Freaake and T. L. Thorp, Rev Sci Instr, 42, 1411 (1971).
4. Richard F. Lacey, Proc. 22nd Freq. Cont. Symp., 545 (1968).
5. L. L. Lewis, M. Feldman and J. C. Bergquist, J. Phys 42, C8-271 (1981).
6. A. DeMarchi, unpublished report dated 22 July 1985.
7. A. DeMarchi, Proc. 40th Freq. Cont. Symp., (1986).
8. A. Brillet, Metrologia, 17, 147 (1981)
V. K. Egorov and V. A. Maslov, Sov Phys Tech Phys, 29, 334 (1984) Jon Shirley, Proc. 39th Freq. Cont. Symp., 22 (1985).

EFFECT OF THE ATOM TRANSIT TIME ON THE FREQUENCY STABILITY OF CESIUM BEAM FREQUENCY STANDARDS

Claude Audoin and Vincent Candelier
Laboratoire de l'Horloge Atomique
Bât. 221 - Université Paris-Sud
91405 ORSAY - France

and

Jacques Vanier
Division de Physique
Conseil National des Recherches
Ottawa, Ontario - Canada K1A 0R6

Abstract

The effect of the transit time of the atoms in cesium beam tube is analyzed, in the case of fast sine-wave and square-wave frequency modulations and in the case of square-wave phase modulation. Analytical expressions are given for the slope of the error signal and for the frequency stability.

The effect of the modulation frequency and of the separation between the second interaction region and the atom detection zone is illustrated for the interaction time distributions encountered when optical methods are implemented for the state selection and the detection of the cesium atoms.

I. Introduction

In passive atomic frequency standards, a modulation of the frequency or of the phase of the interrogation signal is necessary to derive an error signal enabling the frequency control of a quartz crystal oscillator.

Usually, calculation of the error signal and of the frequency stability is made by assuming a quasi-static modulation, in which the modulation frequency is very small compared to the atomic resonance linewidth^{1,2}.

However, in manufactured cesium beam frequency standards, the modulation frequency is equal to approximately 0.3 to 0.4 times the atomic linewidth in the case of sine-wave frequency modulation. Furthermore, one of the other modulation techniques used, the square-wave phase modulation^{3,4}, which is entirely founded on the transient response of the beam tube, is optimized when the modulation frequency is approximately equal to the resonance linewidth. It follows that the assumption of a quasi-static modulation is not sufficient for a correct description of the beam tube response and than an investigation of the influence of the transit time between the two interaction regions and, also, between the second interaction region and the detector is necessary.

In this paper, we give information of the response of a cesium beam tube versus the modulation frequency and calculate the slope of the error signal and the expected white noise limited frequency stability. As it was done in the case of the rubidium cell frequency standard⁵, we shall consider three possible modulation schemes, namely the sine-wave phase (or frequency) modulation, the triangle-wave phase (or square-wave frequency) modulation and the square-wave phase modulation. For the purpose of comparison, it is assumed that the beam tube response is similarly processed i.e. selectively amplified at the modulation frequency and square-wave synchronously detected, whatever the modulation waveform is. The effect of the modulation waveform and of the modulation frequency on the long term frequency stability is addressed in another paper of this issue⁶.

The origin of the linewidth being quite different in the cesium beam tube, the method of derivation will be different and we shall follow the lines of the analysis given by Bell and Babitch⁷. As recalled below, it uses the result of the quasi-static approximation to achieve results valid for modulation frequencies comparable to the line-width. Within this approximation, and considering the central part of the Ramsey pattern, the probability that a transition occurred in the beam tube is given by⁸:

$$P(\tau) = \frac{1}{2} \sin^2 b \tau \{ 1 + \cos [(\omega - \omega_0) T + \phi] \} \quad (1)$$

for monokinetic atoms of velocity v . $\tau = \ell/v$ is the interaction time in each interaction region of length ℓ and $T = \tau L/\ell$ is the transit time between these regions, whose separation is L . b is the representation of the amplitude of the microwave field. It should be noted that the value of b that we are using here differs by a factor of 2 from that introduced by Ramsey. ω is the constant angular frequency of the microwave field and ω_0 is the hyperfine transition angular frequency of the cesium atom. ϕ is the phase of the microwave field in the second interaction region relative to that in the first one.

The observable transit time effect being obviously dependent on the velocity distribution of the detected atoms, we shall perform the appropriate velocity average. In order to illustrate its effect, we shall assume the analytically defined velocity distributions which may be encountered when the cesium atoms are optically pumped and optically detected^{9,12}. However, it should be pointed out that the given analytical results can be used for any distribution of the atom velocity or of the interaction time.

2. Interaction Time Distributions Considered

In optically pumped and optically detected cesium beam tubes, velocity selective effects are produced by the duration of the interaction with the light beams^{13,14}.

In the optical pumping zone, where a population difference is created between the levels $|F=4, m_F=0\rangle$ and $|F=3, m_F=0\rangle$, two cases must be considered. The first one is that of a single wavelength light beam. The transition involved is saturated for most of the interaction times, for the actually achievable light beam power densities. One may thus assume that the population difference achieved does not depend on the velocity. In the second case, a two wavelengths light beam is used to transfer as efficiently as possible the population of the sixteen hyperfine sublevels to a single $m_F=0$ one^{12,15}. However, the saturation of this two photons process requires a larger energy density or a longer interaction time, which may not be achieved in practice. One may then assume, as a first approximation,

that the population difference is proportional to the interaction time with the light beam, which introduces a weighting factor of $1/v$ in the velocity distribution.

In the optical detection zone, two different cases have still to be considered, depending on the detection scheme. A first possibility consists in using the same, single wavelength, light source as in the optical pumping zone. In that case, the saturation condition can be fulfilled as well and the number of collected fluorescence photons does not depend on the velocity. When a different wavelength light source is available, a cycling two level transition is chosen, which yields a number of fluorescence photons proportional to the interaction time and, thus, a weighting factor of $1/v$.

In the following, we shall take into consideration the different cases mentioned by assuming a distribution function of the interaction time τ given by :

$$f_n(\tau) = \frac{C_n}{\tau_0} \left(\frac{\tau_0}{\tau} \right)^{n+2} \exp(-\tau_0^2/\tau^2). \quad (2)$$

For $n = 3$, we have the distribution of interaction time corresponding to the maxwellian beam velocity distribution. For $n = 2$ or 1 , Eq. (2) gives the distribution of interaction time assuming a weighting factor of $1/v$ or $1/v^2$, respectively. $\tau_0 = \lambda/\alpha$ is the interaction time associated with the most probable velocity, α , in the oven⁸. C is a normalization factor given by :

$$\int_0^\infty f_n(\tau) d\tau = 1 \quad (3)$$

We have $C_n = 2$ for $n = 1$ and 3 and $C_n = 4/\sqrt{\pi}$ for $n = 2$.

3. Flux of Detected Atoms

When the optically detected atoms are those which have made the microwave $\Delta F = 1$, $\Delta m_F = 0$ clock transition, their flux is given by :

$$I = I_b + \frac{I_0}{2} \int_0^\infty f(\tau) \sin^2 b\tau (1 + \cos[(\omega - \omega_0)\tau + \phi]) d\tau, \quad (4)$$

where $f(\tau)$ is the interaction distribution which will be made identical to $f_n(\tau)$ when numerical results will be considered. I_0 can be expressed from the source temperature, the beam geometry and the optical pumping efficiency. I_b is a frequency independent background atom flux. It includes : i) a flux representing a possibly incomplete optical pumping, ii) an atom flux equivalent to the stray flux of photons scattered from the detection laser light beam and iii) a flux of atoms whose shot noise is equivalent to the detector noise.

4. Beam Tube Response to Fast Frequency Modulation

In the derivation of Eq(1), it is assumed that the phase of the microwave field does not depend on time during each atom-microwave field interaction. We shall suppose that this remains a valid approximation in the presence of phase modulation^{3,4,7}. This is justified by the fact that we shall consider modulation periods, T_M , which are at most approximately equal to the transit time between the two interaction regions. The ratio τ/T_M is thus small, being at most approximately equal to the ratio λ/L . Square-wave phase modulation implies a discontinuity of the phase and, in that case, a fraction τ/T_M of the atoms experience this step in their interaction with the microwave field. We shall similarly assume that this abrupt phase change does not affect noticeably the validity of the subsequent results. The assumption made amounts to say that the interrogation microwave field has a constant angular frequency, ω_i , and a constant phase during the interaction with the atoms and that the change of phase occurs during the passage from one region of interaction to the other.

The joint effect of the modulation and of the atomic motion is thus to produce an apparent phase difference

between the two oscillatory fields. If t' is the time at which the atoms enter the second interaction region, then $\phi(t')$ is the phase at that time and $\phi(t' - T)$ represents the phase when the atoms entered the first region. The apparent phase difference is thus given by :

$$\phi(t', T) = \phi(t') - \phi(t' - T) \quad (5)$$

We shall ignore any permanent phase shift between the two fields and we shall replace ω by ω_i and ϕ by $\phi(t', T)$ in Eq. (1).

The atoms are detected at the time t , posterior to t' , due to the separation, D , between the second interaction region and the detector. We have

$$t = t' + TD/L, \quad (6)$$

and consequently :

$$\phi(t', T) = \phi[t - TD/L] - \phi[t - T(1 + D/L)]. \quad (7)$$

In Eq. (5) we do not take into consideration the microwave interaction time, according to the assumption $\tau \ll T$ made in the present analysis.

For a small frequency offset, we obtain :

$$I(t) = I_b + \frac{I_0}{2} \int_0^\infty f(\tau) \sin^2 b\tau (1 + \cos[\phi(t', T)]) d\tau - \frac{I_0}{2} (\omega_i - \omega_0) \int_0^\infty \tau f(\tau) \sin^2 b\tau \sin[\phi(t', T)] d\tau. \quad (8)$$

In the cases considered, the spectrum of the periodic phase $\phi(t)$ will contain odd harmonics only. It follows that the Fourier series expansion of $\cos[\phi(t', T)]$ and of $\sin[\phi(t', T)]$ contains only even and odd harmonics, respectively. Therefore, the third term of Eq.(8) provides the error signal, proportional to $(\omega_i - \omega_0)$, after synchronous detection at the angular frequency $\omega_M = 2\pi/T_M$, and low pass filtering.

5. Normalized Slope of the Error Signal

5.1. Definitions

In the following, we shall assume that the fundamental component of the beam tube response is selectively amplified at angular frequency ω_M , so that the fundamental component, $I_1(t)$, of $I(t)$ is observed. Assuming that the offset of the interrogation frequency is small, i.e. that we have $|\omega_i - \omega_0|T \ll 1$, $I_1(t)$ can be written as :

$$I_1(t) = -(\omega_i - \omega_0) I_0 (S \sin \omega_M t + C \cos \omega_M t), \quad (9)$$

where S and C are proportional to the amplitude of the components of $I_1(t)$ which are in phase and in quadrature with the phase waveform, respectively. The phase ϕ of $I_1(t)$ is given by :

$$\tan \phi = C/S \quad (10)$$

Assuming that the demodulation waveform is adjusted in phase with $I_1(t)$, the following error signal is obtained at the low-pass filtered output of the synchronous detector :

$$d = -2K I_0 (\omega_i - \omega_0) (S^2 + C^2)^{1/2} / \pi. \quad (11)$$

The factor $2/\pi$ is related to the synchronous detection by a square waveform and K is a constant which depends on the photon collection efficiency, on the detection amplifier and on the synchronous detector gains. Eq.(11) can be used with the subsequently given equations for S and C to determine the slope of the error signal at $\omega_i = \omega_0$ in the case of any interaction time distribution.

When the interaction time distribution is analytically defined, as in Eq. (2), it is convenient to introduce a dimensionless normalized slope a' , of the error signal. We set :

$$a' = d/KI_0 T_0 (\omega_i - \omega_0), \quad (12)$$

where T_0 is a given value of the transit-time between the two interaction regions. We shall choose $T_0 = L/\alpha = L\tau_0/\ell$.

5.2. Sine-wave phase or frequency modulation

In that case, we have :

$$\phi(t) = \phi_m \sin \omega_M t, \quad (13)$$

where ϕ_m is the amplitude of phase modulation. The instantaneous angular frequency of the microwave field is :

$$\omega(t) = \omega_i + \omega_m \cos \omega_M t, \quad (14)$$

with

$$\omega_m = \phi_m \omega_M. \quad (15)$$

It may be shown that the quantities C and S of Eq. (9) are given by :

$$\begin{pmatrix} C \\ \text{or} \\ S \end{pmatrix} = \int_0^\infty T f(\tau) \sin^2 b\tau J_1 \left[\omega_M T \frac{\sin(\omega_M T/2)}{\omega_M T/2} \right] \begin{pmatrix} \cos \omega_M T(\frac{1}{2} + \frac{D}{L}) \\ \text{or} \\ \sin \omega_M T(\frac{1}{2} + \frac{D}{L}) \end{pmatrix} d\tau. \quad \begin{pmatrix} C \\ \text{or} \\ S \end{pmatrix} = \sin 2\phi_m \int_0^\infty T f(\tau) \sin^2 b\tau \sin \left(\frac{\omega_M T}{2} \right) \begin{pmatrix} \cos \omega_M T(\frac{1}{2} + \frac{D}{L}) \\ \text{or} \\ \sin \omega_M T(\frac{1}{2} + \frac{D}{L}) \end{pmatrix} d\tau \quad (16)$$

In this equation, it is understood that the first order Bessel function in an odd function of its argument.

It follows that the error signal slope and the phase shift ϕ depend on the interaction time distribution, on the normalized microwave level b , on the amplitude of the angular frequency deviation ω_m , on the modulation angular frequency ω_M and on the beam geometry through the ratio D/L . One may retrieve the result of the quasi-static approximation by setting $\omega_M = 0$ in Eq. (16).

5.3. Triangle-wave phase modulation, or square-wave frequency modulation

We consider the following periodic phase modulation waveforms :

$$\begin{cases} \phi(t) = \omega_M t & \text{for } 0 \leq t < T_M/4, \\ \phi(t) = \omega_M (T_M/2 - t) & \text{for } T_M/4 \leq t < 3T_M/4, \\ \phi(t) = \omega_M (t - T_M) & \text{for } 3T_M/4 \leq t < T_M. \end{cases} \quad (17)$$

It corresponds to a square-wave frequency modulation of amplitude ω_m , with frequency jumps occurring at $t = T_M/4 + kT_M/2$, where k is an integer.

The calculation of the quantities C and S implies the knowledge of the fundamental components of the Fourier series expansion of $\sin [\phi(t', T)]$, which are given by :

$$\begin{pmatrix} C_1 \\ \text{or} \\ S_1 \end{pmatrix} = \frac{1}{\pi} \int_0^\infty \sin [\phi(t', T)] \begin{pmatrix} \cos \omega_M t \\ \text{or} \\ \sin \omega_M t \end{pmatrix} d(\omega_M t). \quad (18)$$

We thus have :

$$\begin{pmatrix} C \\ \text{or} \\ S \end{pmatrix} = \int_0^\infty T f(\tau) \sin^2 b\tau \begin{pmatrix} C_1 \\ \text{or} \\ S_1 \end{pmatrix} d\tau. \quad (19)$$

A direct calculation is best suited for the derivation of the analytical expression of a' in the quasi static approximation. Assuming, as before, narrow band amplification of the beam tube response and square-wave synchronous detection, we then have

$$a' = -\frac{4}{\pi^2} \int_0^\infty \frac{\tau}{\tau_0} f(\tau) \sin^2 b\tau \sin \omega_m T d\tau. \quad (20)$$

5.4. Square-wave phase modulation or frequency-impulse modulation

The phase modulation waveform is then given by :

$$\phi(t) = \begin{cases} \phi_m & \text{for } 0 \leq t < T_M/2 \\ -\phi_m & \text{for } T_M/2 \leq t < T_M \end{cases}. \quad (21)$$

The related frequency modulation spectrum is composed of impulse-functions peaking upwards at $t = kT_M$, where k is an integer and downwards at $t = (2k+1)T_M/2$.

This type of modulation, unlike the two others considered, relies entirely upon the transient characteristics of the beam tube. It can be shown that, consequently, the slope of the error signal is optimized for a modulation frequency which is of the order of magnitude of the line-width³.

The analytical expression of $\sin [\phi(t', T)]$ is simple in the present case^{4,7}, and we have :

$$\begin{pmatrix} C \\ \text{or} \\ S \end{pmatrix} = \sin 2\phi_m \int_0^\infty T f(\tau) \sin^2 b\tau \sin \left(\frac{\omega_M T}{2} \right) \begin{pmatrix} \cos \omega_M T(\frac{1}{2} + \frac{D}{L}) \\ \text{or} \\ \sin \omega_M T(\frac{1}{2} + \frac{D}{L}) \end{pmatrix} d\tau \quad (22)$$

Obviously, the value of the amplitude, ϕ_m , of the phase modulation which gives the maximum value of the error signal does not depend on the velocity distribution and it is equal to $\pi/4$.

5.5. Optimization of the error signal

The slope of the error signal characterizes the ability of the beam tube to detect an offset of the interrogation frequency from the hyperfine resonance frequency. Thus, the knowledge of the operating conditions which optimize this quantity is of importance. Furthermore, it will be shown in the next section that, in the presence of a large background and noise equivalent flux of detected atoms or photons, the frequency stability is optimized for the operating conditions considered here.

Since, in general, the value of a' depends on a number of parameters, we shall restrict ourselves to present selected results. Table 1 gives the maximum value of $|a'|$ and the related operating conditions for slow sine-wave and square-wave frequency modulations, for which the value of D/L does not matter. Similarly, Table 2 gives results for square-wave phase modulation, for which the optimum value of the amplitude of the phase step is fixed, equal to $\pi/4$. Figures 1 to 3 show the maximum achievable value of the normalized slope, i.e. for which the relevant operating parameters have been optimized, versus the normalized modulation frequency, for the three different modulation waveform considered, assuming the interaction time distribution for which $n = 2$ in Eq.(2).

The two main general conclusions are the following :
i) when the quasi-static approximation does not hold, the separation between the second interaction region and the atom detection zone has a detrimental effect. It is due to the increasing dispersion of the phase of the response of atoms having different velocities, in this region of the beam tube. The effect is the largest for the smallest values of n , corresponding to the broadest interaction time distributions.

ii) for a given value of D/L and of n , the three modulation waveforms may give, for appropriate values of the operating conditions, very similar values of the slope of the error signal. This value increases when n decreases, in relation with a longer mean value of the interaction time.

Modulation wave-form	Slow sine-wave			Slow square-wave		
	n	1	2	3	1	2
$b \tau_o$	0.79	1.21	1.56	0.77	1.19	1.54
$\omega_m T_o$	0.95	1.45	1.85	0.77	1.19	1.54
$ a' $	0.245	0.220	0.203	0.255	0.233	0.218

Table 1

Slow sine-wave and square-wave frequency modulations. Values of the normalized amplitude of the microwave field, $b \tau_0$, and of the normalized amplitude of frequency modulation, $\omega_m T_0$, for which the fractional slope of the error signal, $|a'|$, has its maximum value. The values of n determine the particular distribution functions of the microwave interaction time considered, according to Eq. (2). The beam tube response is assumed narrow-band filtered at the angular frequency ω_m .

D/L		0	0.25	0.5
n = 1	$\omega_m T_0$	2.21	1.57	1.27
	$b \tau_0$	1.10	0.98	0.95
	$ a' $	0.344	0.248	0.207
n = 2	$\omega_m T_0$	2.78	2.33	2.02
	$b \tau_0$	1.41	1.40	1.39
	$ a' $	0.253	0.213	0.183
n = 3	$\omega_m T_0$	3.30	2.97	2.65
	$b \tau_0$	1.69	1.72	1.72
	$ a' $	0.219	0.196	0.172

Table 2

Square-wave phase modulation. Values of the normalized modulation frequency, $\omega_m T_0$, and of the normalized amplitude of the microwave field, $b \tau_0$, for which the fractional slope of the error signal, $|a'|$, has its maximum value. The values of n determine the particular distribution functions of the microwave interaction time considered according to Eq. (2). The beam tube response is assumed narrow band filtered at the angular frequency ω_m .

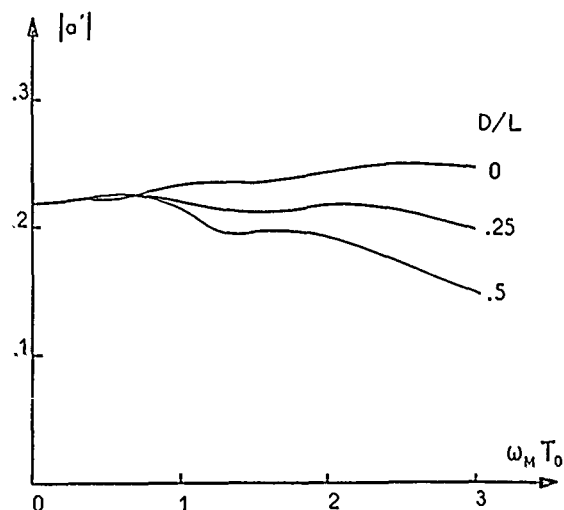


Figure 1

Sine-wave phase modulation. Variation of the maximum value of the fractional slope a' for three different values of the ratio D/L , versus the normalized modulation angular frequency, the beam tube response being selectively amplified at angular frequency ω_m . It is assumed the interaction time distribution $f_2(\tau)$ defined by Eq. (2).

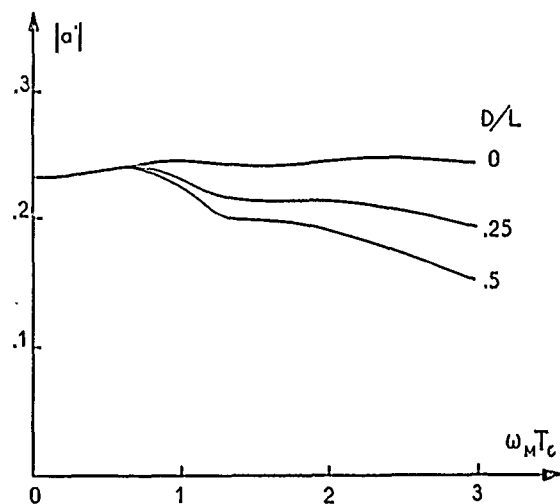


Figure 2

Square-wave frequency modulation. Variation of the maximum value of the fractional slope a' for three different values of the ratio D/L , versus the normalized modulation angular frequency, the beam tube response being selectively amplified at angular frequency ω_m . It is assumed the interaction time distribution $f_2(\tau)$ defined by Eq. (2).

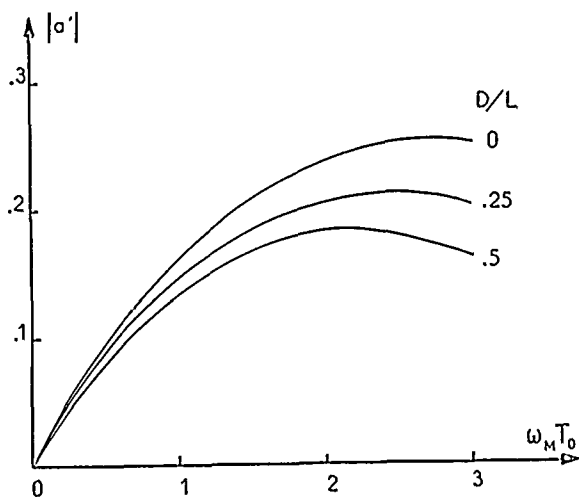


Figure 3

Square-wave phase modulation. Variation of the maximum value of the fractional slope a' for three different values of the ratio D/L , versus the normalized modulation angular frequency, the beam tube response being selectively amplified at angular frequency ω_M . It is assumed the interaction time distribution $f_2(\tau)$ defined by Eq. (2).

6. Frequency Stability

6.1. General result

In passive frequency standards, the function of the frequency control loop of the quartz crystal oscillator is to maintain equal to zero the error signal at the output of the synchronous detector. Consequently any signal due to noise is translated into frequency fluctuations. In the present case, the spurious signal is due to the shot noise of the flux of detected particles and of the detector noise equivalent flux.

If S_n is the one-sided power spectral density of the noise at the low-pass filtered output of the synchronous detector, the one-sided power spectral density of fractional frequency fluctuations is given by :

$$S_y = S_n / (K_d \omega_0)^2 \quad (23)$$

where K_d is the slope of the frequency discriminator composed of the beam tube and the synchronous detector. We have :

$$K_d = d/(\omega_1 - \omega_0) \quad (24)$$

where d is defined by Eq. (11). S_n characterizes the intrinsic frequency stability of the cesium beam tube in the frequency domain. The related measure of frequency stability in the time domain, the two-sample variance, is given by $\sigma^2(\tau) = S_y / 2\tau$. The quantities S_n and $\sigma^2(\tau)$ determine the frequency stability of the association of the cesium beam tube and of the quartz crystal oscillator for sampling times, τ , larger than the time constant of the frequency loop¹.

It may be shown that the one-sided power spectral density of the noise at the low pass filtered output of the synchronous detector is given by :

$$S_n = 2(8/\pi^2) K^2 F (I_{dc} + I_b) \quad (25)$$

The factor $8/\pi^2$ comes from the effect of the square-wave demodulation on the input noise, assumed selectively amplified at the modulation frequency (see Appendix A). The factor K has the same meaning as above. F is the noise factor of the particle detector. I_{dc} is the time average value of the flux modulated by the atomic line interrogation (see Appendix B).

We thus have, from Eqs. (11) and (23) to (25) :

$$S_y = 4 F (I_{dc} + I_b) / I_0^2 (S^2 + C^2) \omega_0^2 \quad (26)$$

This equation is valid in general, for any interaction time distribution. When it is convenient to introduce the normalized slope a' , it becomes :

$$S_y = 16 F (I_{dc} + I_b) / (\pi a' I_0 T_0 \omega_0)^2 \quad (27)$$

6.2. Case of a large background and noise equivalent flux

When the sum, I_b , of the background and added noise equivalent flux is large compared to the line height, which is defined as the difference between the peak and the valley fluxes of detected atoms, the power spectral density of fractional frequency fluctuations does not depend on the modulation process applied to interrogate the microwave transition. The frequency stability is thus optimized when $(S^2 + C^2)$ or $|a'|$ show their maximum possible values. They have been considered in Section 5 for the interaction time distributions defined by Eq. (2).

6.3. Case of a small background and noise equivalent flux

The other, more satisfactory limiting case, is that where I_b is much smaller than the line-height. It is achieved when no spurious flux contributes to shot noise. In that case, we thus have to consider the time average value of the periodically modulated flux, which depends on the modulation waveform and on the modulation frequency.

This time average value is proportional to I_0 , and for the interaction time distributions defined by Eq.(2), we define the dimensionless parameter H such as :

$$S_y = F H / \omega_0^2 T_0^2 I_0 \quad (28)$$

It depends on the interaction time distribution, on the amplitude of the microwave field, on the modulation waveform, on the modulation amplitude and on its frequency. It enables to compare the expected optimum frequency stability for the different distribution functions and modulation schemes considered.

6.3.1. Slow modulation. We have :

$$I_{dc} = \frac{I_0}{2} \int_0^\infty f(\tau) \sin^2 b\tau [1 + J_0(\omega_m T)] d\tau \quad (29)$$

for slow sine-wave frequency modulation, and

$$I_{dc} = \frac{I_0}{2} \int_0^\infty f(\tau) \sin^2 b\tau (1 + \cos \omega_m T) d\tau \quad (30)$$

for slow square-wave frequency modulation.

With the interaction time distribution given by Eq. (2), the related values of the parameter H are respectively given by¹⁷ :

$$H = 2 \frac{\int_0^\infty f(\tau) \sin^2 b\tau [1 + J_0(\omega_m T)] d\tau}{\left\{ \int_0^\infty \frac{\tau}{T_0} f(\tau) \sin^2 b\tau J_1(\omega_m T) d\tau \right\}^2} \quad (31)$$

and

$$H = \frac{\pi^2}{2} \frac{\int_0^\infty f(\tau) \sin^2 b\tau (1 + \cos \omega_m T) d\tau}{\left\{ \int_0^\infty \frac{\tau}{T_0} f(\tau) \sin^2 b\tau \sin \omega_m T d\tau \right\}^2} \quad (32)$$

Table 3 gives, for $n = 1, 2$ and 3 , the operating conditions for which the value of the parameter H shows a minimum, corresponding to the best possible frequency stability. It shows that slow square-wave frequency modulation has a little better frequency stability capability and that the smallest values of n favors the frequency stability, in relation with a longer mean value of the interaction time, and, thus of a larger line quality factor.

Modulation wave-form	Slow sine-wave			Slow square-wave		
	n	1	2	3	1	2
$b\tau_o$	0.48	1.01	1.39	0.53	1.05	1.44
$\omega_m T_o$	0.78	1.52	2.04	0.80	1.48	1.98
H	9.7	15.4	19.8	7.2	10.6	13.0

Table 3

Slow sine-wave and square-wave frequency modulations and narrow band amplification of the beam tube response. Values of the operating parameters, for which the parameter H has its minimum value. The values of n determine the particular interaction time distribution functions, according to Eq. (2).

6.3.2. Fast frequency modulation. According to Eq. (8), the mean value of the detected particle flux is given by :

$$I_{dc} = \frac{I_0}{2} \int_0^\infty f(\tau) \sin^2 b\tau (1 + C_0) d\tau, \quad (33)$$

with

$$C_0 = \frac{1}{2\pi} \int_0^{2\pi} \cos [\phi(t', T)] d(\omega_m t). \quad (34)$$

When the beam tube response is narrow-band filtered at angular frequency ω_m , we thus have :

$$H = \frac{8}{\pi^2} \frac{\int_0^\infty f(\tau) \sin^2 b\tau (1 + C_0) d\tau}{(a')^2}. \quad (35)$$

Figures 4 to 6 show the minimum achievable value of the parameter H , i.e. that for which the relevant operating parameters have been optimized, versus the normalized modulation frequency, for the three different modulation waveforms considered, assuming the interaction time distribution for which $n = 2$ in Eq. (2). Table 4 gives, in the case of square-wave phase modulation, for $n = 1, 2$ and 3 and for $D/L = 0, 0.25$ and 0.5 , the value of the operating parameters which optimize the frequency stability.

The two main conclusions stated at the end of Section 5.5 still apply here. An additional general result is that the value of the operating parameters are different, according as the Ramsey probability, the slope of the error signal, or the frequency stability (in the limiting case of a small background and noise equivalent flux) are optimized.

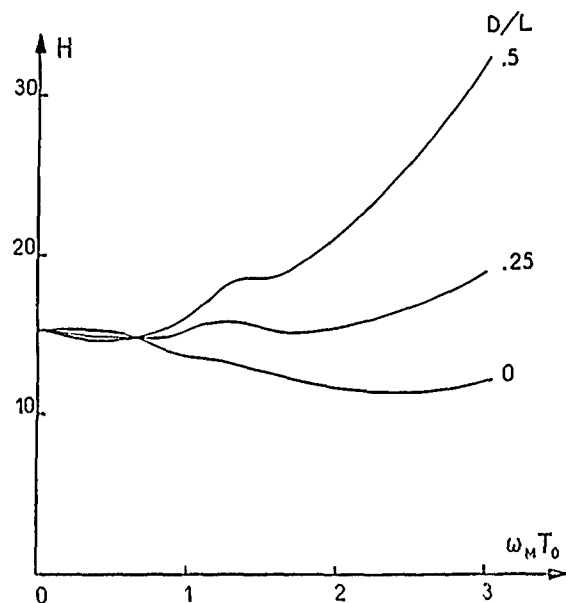


Figure 4

Sine-wave phase modulation. Variation of the minimum value of the parameter H , which specifies the frequency stability, for three different values of the ratio D/L , versus the normalized modulation frequency, the beam tube response being selectively amplified at angular frequency ω_m . It is assumed the interaction time distribution $f_2(\tau)$ defined by Eq. (2).

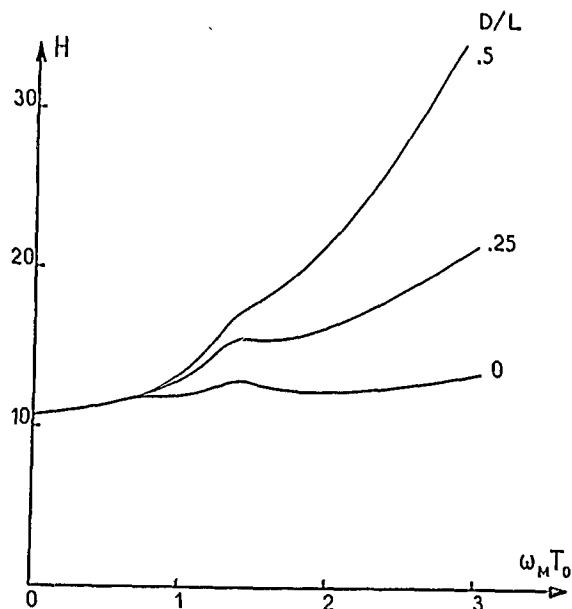


Figure 5

Square-wave frequency modulation. Variation of the minimum value of the parameter H , which specified the frequency stability, for three different values of the ratio D/L , versus the normalized modulation frequency, the beam tube response being selectively amplified at angular frequency ω_m . It is assumed the interaction time distribution $f_2(\tau)$ defined by Eq. (2).

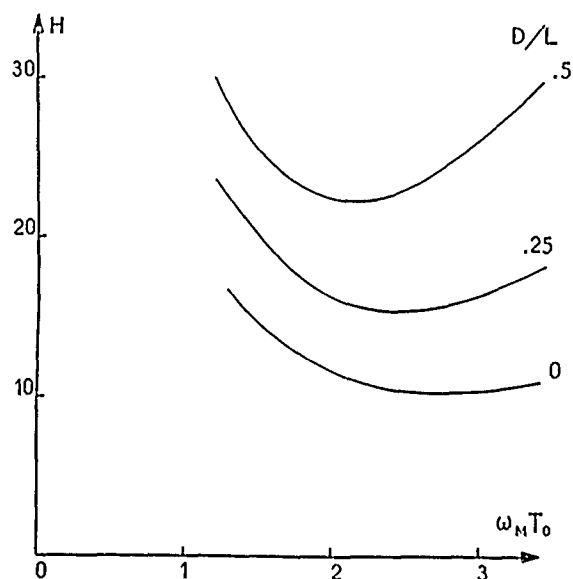


Figure 6

Square-wave phase modulation. Variation of the minimum value of the parameter H , which specifies the frequency stability, for three different values of the ratio D/L , versus the normalized modulation frequency, the beam tube response being selectively amplified at angular frequency ω_M . It is assumed the interaction time distribution $f_2(\tau)$ defined by Eq. (2).

D/L		0	0.25	0.5
n = 1	$\omega_M T_0$	2.21	1.39	1.17
	$b \tau_0$	1.11	0.87	0.87
	ϕ_m	0.94	0.90	0.87
	H	5.3	10.8	16.7
n = 2	$\omega_M T_0$	2.77	2.47	2.18
	$b \tau_0$	1.38	1.37	1.34
	ϕ_m	0.95	0.93	0.92
	H	10.3	15.4	22.3
n = 3	$\omega_M T_0$	3.41	3.23	2.96
	$b \tau_0$	1.69	1.74	1.70
	ϕ_m	0.96	0.95	0.94
	H	14.1	18.5	25.1

Table 4

Square-wave modulation and narrow band amplification of the beam tube response. Values of the operating parameters for which the parameter H has its minimum value for different values of the ratio D/L . The values of n determine the particular interaction time distribution functions, according to Eq. (2).

7. Conclusion

We have given analytical equations allowing to calculate the effect of the modulation frequency on the slope of the error signal and on the expected white noise limited frequency stability, for three commonly used modulation waveforms. Numerical results have been given in the case of the interaction time distributions which may be encountered when optical methods are applied for the state selection and the detection of cesium atoms.

Acknowledgements

The authors are grateful to Henri Clergeot for helpful and stimulating discussions on the subjects of the noise synchronous detection and of the shot noise of a periodically modulated particle flux.

Appendix A. Noise and Synchronous Detection

In a synchronous detector, a noise process, $x(t)$, assumed wide sense stationary, with a zero mean value, is multiplied by a periodic signal $g'(t)$, and the result of this processing $z(t) = x(t)g'(t)$ is low-pass filtered. We shall consider square-wave demodulation, with

$$\begin{aligned} g'(t) &= +1 \text{ for } 0 \leq t < T_M/2 \\ g'(t) &= -1 \text{ for } T_M/2 \leq t < T_M \end{aligned} \quad (\text{A-1})$$

where T_M is the period of the demodulation waveform.

It may be shown easily that the two-sided power spectral density of $z(t)$ is given by :

$$S_z'(f) = \frac{4}{\pi^2} \sum_{p=0}^{\infty} \frac{1}{(2p+1)^2} \{ S_x'[f - (2p+1)f_M] + S_x'[f + (2p+1)f_M] \} \quad (\text{A-2})$$

where S_x' is the two-sided power spectral density of $x(t)$.

The signal $z(t)$ being low pass filtered, we consider the low frequency component of the output signal, at frequencies such as $f \ll f_M$, where f_M is the frequency of $g'(t)$. The process $x(t)$ being a real function of time, we have, for $f \ll f_M$:

$$S_x'[f - (2p+1)f_M] \approx S_x'[f + (2p+1)f_M] \quad (\text{A-3})$$

provided that S_x' varies smoothly around frequencies $\pm (2p+1)f_M$. It follows that, for $f \ll f_M$, the one sided power spectral density of $z(t)$ is given by :

$$S_z(f \ll f_M) \approx \frac{8}{\pi^2} \sum_{p=0}^{\infty} \frac{1}{(2p+1)^2} S_x[(2p+1)f_M] \quad (\text{A-4})$$

where S_x is the one-sided power spectral density $x(t)$. This equation shows that in the frequency range $0 < f \ll f_M$, $z(t)$ is white noise.

When $x(t)$ is selectively around frequency f_M or, more precisely, when its frequency components at frequencies $(2p+1)f_M$, with $p > 0$ are filtered out, the preceding equation becomes :

$$S_z(f \ll f_M) \approx \frac{8}{\pi^2} S_x(f_M) \quad (\text{A-5})$$

Appendix B. Shot Noise of a Periodically Modulated Particle Flux

The derivation of the power spectral density of shot noise of a periodically modulated particle flux follows closely that given by Papoulis for an uniform particle flux¹⁷. The results given in the quoted reference can be easily generalized to arrive at the following equation for the autocorrelation $R_y(t_1, t_2)$ of the Poisson increments $y(t)$ (see reference¹⁷ for definitions and notations) :

$$R_y(t_1, t_2) = \lambda(t_1) \lambda(t_2), \text{ for } |t_1 - t_2| > \varepsilon, \quad (\text{B-1})$$

and

$$R_y(t_1, t_2) = \lambda(t_1) \lambda(t_2) + \frac{1}{\varepsilon} \lambda(t_2) - \frac{1}{\varepsilon} |t_1 - t_2| \lambda(t_2), \quad (\text{B-2})$$

for $0 < |t_1, t_2| < \varepsilon$. $\lambda(t)$ is the value of the particle flux assuming no shot noise and ε is a very small time interval.

The density $\lambda(t)$ being assumed periodic, we have :

$$\lambda(t) = \lambda_0 + \sum_{n=1}^{\infty} \lambda_n \sin(n\omega_M t + \phi_n) \quad (\text{B-3})$$

where ω_M is the angular frequency of the fundamental component of $\lambda(t)$ and λ_0 , λ_n and ϕ_n are constants. We have the following time average values :

$$\langle \lambda(t_2) \rangle = \lambda_0 \quad (\text{B-4})$$

and

$$\langle \lambda(t_1) \lambda(t_2) \rangle = \lambda_0^2 + \frac{1}{2} \sum_{n=1}^{\infty} \lambda_n^2 \cos n\omega_M(t_1 - t_2) \quad (\text{B-5})$$

Therefore, the time average of the autocorrelation $R_y(t_1, t_2)$ depends on $t_1 - t_2 = \tau$ only and may be identified with the autocorrelation function, $R_y(\tau)$ of $y(t)$. We have

$$R_y(\tau) = \lambda_0^2 + \frac{1}{2} \sum_{n=1}^{\infty} \lambda_n^2 \cos n\omega_M \tau, \text{ for } |\tau| > \varepsilon, \quad (\text{B-6})$$

$$R_y(\tau) = \lambda_0^2 + \frac{1}{2} \sum_{n=1}^{\infty} \lambda_n^2 \cos n\omega_M \tau + \frac{\lambda_0}{\varepsilon} - \frac{\lambda_0}{\varepsilon^2} |\tau|, \quad (\text{B-7})$$

for $0 < |\tau| < \varepsilon$.

The Poisson impulses are obtained by letting ε tends to zero. We obtain :

$$R_z(\tau) = \lambda_0^2 + \frac{1}{2} \sum_{n=1}^{\infty} \lambda_n^2 \cos n\omega_M \tau + \lambda_0 \delta(\tau) \quad (\text{B-8})$$

for any τ .

The two sided power spectral density of the process $z(t)$ is then given by :

$$S_z(f) = \lambda_0 + \lambda_0^2 \delta(f) + \frac{1}{4} \sum_{n=1}^{\infty} \lambda_n^2 [\delta(f - n f_M) + \delta(f + n f_M)] \quad (\text{B-9})$$

where f_M is the frequency of modulation. This equation shows that the power spectrum of $z(t)$ is composed of two parts. The first one is frequency-independent and is equal to λ_0 , the average value of the periodic density $\lambda(t)$. The second part is a series of bright lines at frequencies 0 and $\pm n f_M$. It is merely the power spectrum of the periodic density $\lambda(t)$.

Taking into account the fact that, in the problem treated in the main text, the fundamental component of the modulated beam flux is absent, thanks to the frequency control loop and that synchronous detection gives information on the components of the input signal around frequencies $\pm (2p + 1)f_M$ only, Eq. (25) follows.

References

- 1 L.S. Cutler and C.L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards". PROC. of the IEEE 54, n° 2, Feb. 1966, pp. 136-154
- 2 R.F. Lacey, A.L. Helgesson and J.H. Holloway, "Short-Term Stability of Passive Atomic Frequency Standards". Proc. of the IEEE 54, n° 2, Feb. 1966, pp. 170-176
- 3 R.S. Badessa, V.J. Bates and C.L. Searle, "Frequency Impulse Modulation as a Means of Attaining Accuracy in Cesium Atomic Clocks". IEEE Trans. on Instr. and Meas. IM-13, Dec. 1964, pp. 175-180
- 4 S.G. Kukolich and K.W. Billman, "Square-Wave Phase Modulation in Ramsey-Type Molecular Beam Resonance Experiments". Journal of Applied Physics 38, n° 4, March 1976, pp. 1826-1830
- 5 C. Audoin, J. Viennet, N. Cyr and J. Vanier, "Influence of Modulation Frequency in Rubidium Cell Frequency Standards". Proc. of the 14th Annual Precise Time and Time Interval. Applications and Planning Meeting, Greenbelt, Maryland, 1982, pp. 87-110
- 6 A. de Marchi, G.D. Rovera and A. Premoli, "Effects of Servo Loop Modulation in Atomic Beam Frequency Standards Employing a Ramsey Cavity". IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control. To be published.
- 7 C.M. Bell and D. Babitch, "A System Analysis of the Cesium Beam Atomic Clock". IEEE Trans. on Instr. and Meas. IM-17, n° 2, June 1968, pp. 155-166
- 8 N.F. Ramsey in "Molecular Beams", Oxford University Press, London 1956
- 9 A. Kastler, "Quelques suggestions concernant la production optique et la détection optique d'une inégalité de populations des niveaux de quantification spatiale des atomes. Application à l'expérience de Stern et Gerlach et à la résonance magnétique". Journal de Physique et le Radium 11 (1950), pp. 255-265
- 10 M. Arditi and J.L. Picqué, "A Cesium Beam Atomic Clock Using Laser Optical Pumping. Preliminary Tests". Journal de Physique - Lettres 41, Aug. 1980, pp. L379-L381
- 11 L.L. Lewis and M. Feldman, "Optical Pumping by Lasers in Atomic Frequency Standards". Proc. of the 35th Annual Frequency Control, Philadelphia, 1981, pp. 612-624
- 12 G. Avila, E. de Clercq, M. de Labachellerie and P. Cézé, "Microwave Ramsey Resonances from a Laser Diode Optically Pumped Cesium Beam Resonator". IEEE Trans. on Instr. and Meas. IM-34, n° 2, June 1985, pp. 139-143
- 13 P. Cézé, G. Avila, E. de Clercq, M. de Labachellerie and M. Têtu, "Results on a Laser Diode Optically Pumped Cesium Beam". Proc. of the 38th Annual Frequency Control Symposium, Philadelphia, 1984, pp. 452-457
- 14 A. Derbyshire, R.E. Drullinger, M. Feldman, D.J. Glaze, D. Hilliard, D.A. Howe, L.L. Lewis and J.H. Shirley, "Optically Pumped Small Cesium Beam Standards : A Status Report". Proc. of the 39th Annual Frequency Control Symposium, Philadelphia, 1985, pp. 18-21
- 15 L.L. Lewis, "Limitations of Atomic Beam Frequency Standards". Prog. Quant. Elect. 8, 1984, pp. 153-159
- 16 J.A. Barnes, A.R. Chi, L.S. Cutler, D.J. Healy, D.B. Leeson, T.E. Mc Gunigal, J.A. Mullen, W.L. Smith, R.L. Sydnor, R.F.C. Vessot and G.M.R. Winkler, "Characterization of Frequency Stability". IEEE Trans. on Instr. and Meas. IM-20, n° 2, May 1971, pp. 105-120
- 17 The term J_2 should be dropped in Eqs. (61) and (62) of reference 1 and in Eq. (35) of reference 2.

¹⁸ A. Papoulis in "Probability, Random Variables and Stochastic Processes". Mc Graw Hill Kogakusha, Ltd. (1965)

A novel cavity design for Minimization
of distributed phase shift in atomic beam frequency standards

A. De Marchi

Istituto Elettrotecnico Nazionale "Galileo Ferraris" - Torino

Abstract

Distributed phase shift may be the last important accuracy problem at the level of parts in 10^{14} for future Cesium beam primary standards. Its cause is the spatial phase variation in the microwave cavity produced by the Poynting vector which carries the power dissipated in the walls.

In this paper a new cavity design is presented which minimizes this problem by the use of two equal counterrunning waves in a ring, so that a parabolic minimum of the phase is obtained.

The effects of asymmetries in the structure are studied.

1.- Introduction

Distributed phase shift has been often indicated as an important source of inaccuracy in primary Cs beam frequency standards [1].

Although its role has been recently put in a different perspective [2] by a reevaluation of the effects produced by neighboring transitions [3], and by a critical review of beam reversal phase-shift measurements taken over the years on NBS-6 [4], it is still desirable in planning for the next generation, possibly the last and ultimate, of Cs beam frequency standards, to think of ways to reduce such effect.

In fact, even for the most sensible beam positioning, grazing the end short circuit of the microwave cavity, where the phase slope is the least for the magnetic field, frequency shifts of the order of 1×10^{-13} per mm of instability or repositioning imprecision of the beam gravity center are to be expected for primary Cesium beam standards [2].

Furthermore the beam is most often positioned at $\lambda/2$ from the short for signal level reasons (beam cross section) and the slope there is a factor of 3 greater.

Leaving alone the latter, even the former appears unacceptable in view of an overall accuracy goal of 10^{-14} . A different cavity design is needed to overcome this difficulty, being impractical the only two existing proposals of thin beams and superconducting cavities [5].

Since the phase gradient inside the cavity is caused by the residual Poynting vector corresponding to the energy which is dissipated in the walls, the only solution to the problem appears to be interrogating the beam with two equal counterrunning waves, transversal to the beam. In the plane where these two waves have equal intensity the Poynting vector is null and the phase is stationary.

In this paper an implementation of this idea is presented and a microwave structure is proposed in which the position of the stationary phase point is guaranteed by symmetry. The uncertainties introduced by possible imperfections in the symmetry of the cavity are also studied.

2.- Proposed new cavity

Keeping in mind that a variety of similar cavities may be designed, the solution here proposed is to substitute the end sections of the usual Ramsey cavity with a ring cavity each side according to the sketch of Fig. 1. The whole structure can be built for Cesium in standard X band waveguide, rotated from the usual position of commercial standards, so that the rf magnetic field H be parallel to the beam. The advantage of such an arrangement are underlined in [6].

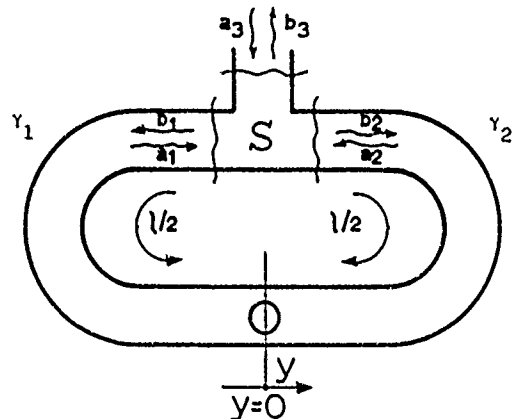


Fig. 1 - Sketch of the studied ring cavity with symbol definitions.

The specific goals for which this cavity has been designed are: easy realization of good symmetry, and good field homogeneity at the beam passing hole. The latter requirement imposes a straight section of waveguide about the hole and the feeding E-plane Tee, the former asks for smooth transitions between straight and curved sections and, as we will show, takes advantage of an E-plane Tee.

The ring admits at resonance two orthogonal resonant modes (and any combination thereof), which are degenerate at the limit of infinitesimal coupling and get frequency separated by selective coupling of one of them. They can be studied as TE_{01n} modes in the waveguide ring or alternatively as TE_{n01} modes in the half wavelength long coaxial cavity. The latter perspective is more precise for cavity dimensioning, the former will be used here for symmetry considerations.

The scattering matrix formalism will be used throughout, with the Tee feed defined by

$$b_i = S_{ij} a_j \quad (1)$$

and the Tee will be considered lossless ($SS^* = 1$). The losses of the two arms will be considered in the real part of the propagation constant $\gamma_i = \alpha_i + j\beta_i$, where γ_i is the average propagation constant of one arm; $\gamma = (\gamma_1 + \gamma_2)/2$ will be the average propagation constant of the ring and γ_d , γ_1 , γ_2 will be the arm asymmetry.

It is for the ring:

$$a_1 = e^{-\gamma d} b_2; \quad a_2 = e^{-\gamma d} b_1 \quad (2)$$

3.- The symmetry problem

It is soon realized that in the proposed structure a difference γ_d in the propagation constant of the two arms, particularly its real part α_d , will move the stationary phase point from the geometrically symmetric point $y = 0$ because a travelling wave will then be present there. It is also true that a perfectly symmetric Tee feed will selectively excite only the desired mode, but an asymmetry of the feed will result in some excitation of the orthogonal unwanted mode [7], inducing a shift of the stationary phase point. (It is known that in an optical ring cavity the position of the electromagnetic nodes cannot be predicted [8]. This may cause some concern).

In the following the effects of asymmetries are studied for the relevant transverse magnetic field H with reference to the structure of Fig. 1.

The phase of $H(y)$ at a distance y from the geometrically symmetric point $y = 0$, referred to the phase of $H(0)$, is the argument of the complex number

$$\frac{H(y)}{H(0)} = \frac{b_1(y) - a_1(y)}{b_1(0) - a_1(0)} = \frac{\text{ch}(\gamma y + \gamma_d l/4 - \psi)}{\text{ch}(\gamma_d l/4 - \psi)} \quad (3)$$

where

$$\psi = \frac{1}{2} \ln(-b_1/b_2) = \frac{1}{2} \ln(1 + 2\epsilon_b) = \epsilon'_b + j\epsilon''_b \quad (4)$$

In fact, since the junction is an E-plane Tee, it is $b_1 = -b_2$ at symmetry [9] when signs consistent with (2) are used at the symmetric ports. Small asymmetries have been assumed for the linearization of (4).

The stationary phase point is at the minimum of the argument of (3), that is at:

$$y = -\frac{l}{8} \left(\frac{\alpha_d}{\alpha} + \frac{\beta_d}{\beta} \right) + \frac{1}{2} \left(\frac{\epsilon'_b}{\alpha} + \frac{\epsilon''_b}{\beta} \right) \quad (5)$$

A first comment to be made is that symmetry requirements on γ are looser for smaller rings. In particular the shift y produced by a difference α_d in average attenuation in the two arms is only 0.1 mm for

a two wavelength ring if $\alpha_d/\alpha = 10^{-2}$, while it would become 1 mm for a 20 wavelength ring. The requirement on β_d is less difficult to fulfill.

On the other hand also for the part due to Tee asymmetry the real part of ϵ_b is most important. It will be illustrated under which circumstances the latter can be expected to be small.

4.- Asymmetry of the feed and the excitation of the orthogonal mode

The deviation from 1 of $-b_1/b_2$, that is ϵ_b , must be calculated in the actual working conditions, that is when the ring is at resonance for the desired mode.

The expressions of b_i as a function of the scattering matrix of the Tee are easily derived with the help of eq. (2)

$$\begin{cases} b_1 = \frac{S_{13} (1 - G_1 e^{-\gamma d})}{(S_{12}^2 - S_{11} S_{22}) e^{-2\gamma d} - 2S_{12} e^{-\gamma d} + 1} a_3 \\ b_2 = \frac{S_{23} (1 - G_2 e^{-\gamma d})}{(S_{12}^2 - S_{11} S_{22}) e^{-2\gamma d} - 2S_{12} e^{-\gamma d} + 1} a_3 \end{cases} \quad (6a)$$

$$G_1 = S_{12} - S_{11} S_{23}/S_{13}; \quad G_2 = S_{12} - S_{22} S_{13}/S_{23} \quad (6b)$$

For a perfectly symmetric Tee these simplify to become

$$b_1 = -b_2 = \frac{T}{1 - K e^{-\gamma d}} a_3 \quad (7a)$$

$$G_1 = G_2 = R + M = e^{j\delta} \quad (7b)$$

where the scattering matrix of the symmetric Tee has been defined as

$$S = \begin{pmatrix} R & M & T \\ M & R & -T \\ T & -T & K \end{pmatrix} \quad (8)$$

and terminal planes at the three ports have been chosen so that T and K are real. This choice is made both for simplicity and because it provides the best input coupling at resonance.

It must be pointed out that imposing the losslessness condition $SS^* = 1$ the following equations are obtained

$$\begin{cases} |R|^2 + |M|^2 + T^2 = 1 \\ 2T^2 + K^2 = 1 \\ 2\text{Re}(RM^*) = T^2 \\ M - R = K \end{cases} \quad (9)$$

and therefore of the 6 parameters defining S (two real and two complex) only two are independent. Since from (9) we get $|R+M| = 1$ we can write (7b) and the two parameters T and δ will be used in the following.

In order to define now the structure of changes occurring in the scattering matrix for small deviations from symmetry, the following assumptions will be made on S and \bar{S} (S-matrix of the deformed Tee):

$$\begin{cases} \bar{S} = S + E & \text{with } EE^* \approx 0 \\ SS^* = \bar{S}\bar{S}^* = 1 \\ U\bar{S}U = S - E \end{cases} \quad (10) \quad (11) \quad (12)$$

The first implies that asymmetries are small, the second that the Tee is still lossless after deformations, and the third that exchanging ports 1 and 2 the effects of the asymmetries are inverted.

Here U is the operator that exchanges ports 1 and 2, which can be represented in matrix form by

$$U = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad (13)$$

From (12) we get $UEU = -E$, which imposes

$$E = \begin{pmatrix} \epsilon_R & 0 & \epsilon_T \\ 0 & -\epsilon_R & \epsilon_T \\ \epsilon_T & \epsilon_T & 0 \end{pmatrix} \quad (14)$$

It is now clear that E is the scattering matrix of the Tee which defines the excitation of the unwanted orthogonal mode, for which ϵ_T gives the transmission parameter.

The relation between ϵ_R and ϵ_T can be obtained from (10) and (11), which yield

$$SE^* + ES^* = 0 \quad (15)$$

A set of 4 (real) equations in 4 (real) unknown is obtained, but only two of them are independent and the relation is found

$$\epsilon_R = -\frac{2}{T} [\operatorname{Re}(Mc_T^*) + j\operatorname{Im}(R\epsilon_T^*)] \quad (16)$$

With the help of (14) and (16) we can now rewrite eq. (6a) to find $-b_1/b_2$ for an asymmetric Tee. Notice that the denominator of eq. (6a) is independent of small asymmetries.

We find:

$$\begin{cases} G_1 = (R+M)(1-2\epsilon_T^*/T) \\ G_2 = (R+M)(1+2\epsilon_T^*/T) \end{cases} \quad (17)$$

from which:

$$\begin{cases} -\frac{b_1}{b_2} = \left(1 + 2\frac{\epsilon_T}{T}\right) \left(1 - 2\frac{\epsilon_T^* F}{T}\right) \\ F = 1 - j\operatorname{ctg}(\delta + j\gamma\ell)/2 \end{cases} \quad (18)$$

and therefore

$$\epsilon_b = j \left(\frac{\epsilon_T''}{T} + \frac{\epsilon_T'}{T} \operatorname{ctg}(\delta + j\gamma\ell)/2 \right) \quad (19)$$

This is the desired expression of ϵ_b induced by a small asymmetry of the Tee, which tells us how much of the orthogonal mode is excited when the Tee is not perfectly symmetric. Since ϵ_b is clearly dependent on frequency we must study now the resonance conditions, and then (19) can be used to find with (5) the displacement of the stationary phase point.

5.- Resonance conditions

To find the resonance condition for the desired mode we impose that the reflection coefficient Γ_3 at port 3 be real. This is satisfied as usual for a frequency which depends on the selected reference plane at port 3. It can be shown that the energy transfer (coupling) at resonance also depends on such plane, and that it is maximized for the reference plane which makes K real. We will call ν_0 the resonance frequency in this case and β_0 the corresponding propagation constant. It is at symmetry:

$$\Gamma_3 = \frac{b_3}{a_3} = T e^{-\gamma\ell} \left(\frac{b_2}{a_3} - \frac{b_1}{a_3} \right) + K = \frac{K - e^{-\gamma\ell}}{1 - K e^{-\gamma\ell}} \quad (20)$$

where T has also been assumed real. It is easily seen that Γ_3 is real for $\sin \beta_0 \ell = 0$, in which case

$$\Gamma_3 = -1 + \frac{1+K}{1-K} \alpha\ell \quad (21)$$

It has been taken into account here that $\cos \beta_0 \ell = 1$ because the reference planes for T real are only 1 mm apart, as shown in the appendix, and an integer number n of wavelengths must then close the ring at resonance. The quality factor Q_0 of this resonance will be practically equal to the external Q, since the resonator is strongly coupled to the outside world, and can be between a few hundred and a thousand. To find now the resonance of the orthogonal unwanted mode the frequency behaviour of (19) must be studied. Introducing the angle ϕ :

$$\phi = \frac{1}{2} (\delta - (\beta - \beta_0)\ell) \quad (22)$$

we can write:

$$j\operatorname{ctg}(\delta + j\gamma\ell)/2 = \frac{j \sin \phi \cos \phi + \alpha\ell/2}{\sin^2 \phi + (\alpha\ell/2)^2} \quad (23)$$

which describes a resonance around $\phi = 0$, at frequency ν_1 and with quality factor Q_1 given by:

$$(\nu_1 - \nu_0)/\nu = \delta/2\pi n \quad (24)$$

$$Q_1 \approx 2\pi/\alpha\lambda_g \quad (25)$$

It appears from (25) that Q_1 is very high, due to the small coupling, and from (24) that it is not until $n \approx Q_1 |\delta|/2\pi$ that we run the risk of exciting the unwanted mode with measurable power.

This is easily avoided for a ring of a few wavelengths.

However, because in the expression of the shift y of the stationary phase point the real part of ϵ_p is divided by α , it is still important to include the undesired orthogonal mode in its calculation.

6.- Shift of the stationary phase point

From (5) and (19), with the help of (23) and $\phi \neq 0$ (out of resonance for the undesired mode) we can write:

$$\frac{y}{\lambda_g} \approx \frac{1}{4\pi} \frac{\epsilon''_T}{T} + \frac{n}{4 \sin^2 \phi} \frac{\epsilon'_T}{T} \quad (26)$$

valid for $n > 1$ and $y/\lambda_g \ll 1$. It is clear that in an optical ring the node position comes out totally undefined by the feed [8]. For a ring of a few wavelengths however it is not so. The value of δ is important because it gives a measure of the distance of the two modes; a small $|\delta|$ makes the second term of (26) critical if ϵ'_T is of any importance.

In an open Tee we calculate in the appendix $\delta = -0.14$ and for $n = 2$ this imposes $\epsilon'_T/T < 10^{-4}$ for $y/\lambda_g < 10^{-2}$. This may be an excessive requirement if ϵ'_T/T depends linearly on the geometrical asymmetry, as it would presumably be for an H plane Tee.

Although exact computation of ϵ_T/T for an E-plane Tee is a difficult electromagnetic problem, which will not be tackled here, symmetry considerations can prove that the only type of Tee asymmetry relevant to our discussion is a tilt ϑ from perpendicularity of the feeding arm in the plane E, and an inspection of the E field patterns in the junction (Fig. 2) suggests that the phase of the transmitted fields be linearly dependent on ϑ , while their amplitude on a higher power of ϑ , at least the second, more likely the third. Assuming as a conservative reasonable guess that $\epsilon'_T/T \approx (\partial \vartheta b/2)^2 \approx \vartheta^2/2$ and $\epsilon''_T/\vartheta \approx (\partial \vartheta b/2) \approx 0.7 \vartheta$, we find for the open Tee that $y/\lambda_g < 10^{-2}$ if $\vartheta < 1.4 \times 10^{-2}$. This is a reasonable specification.

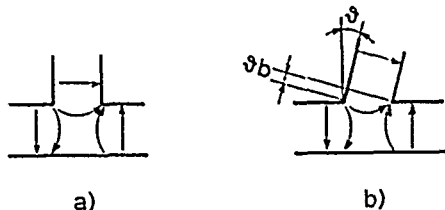


Fig. 2- E-field patterns in a symmetric (a) and asymmetric (b) E-plane Tee. It appears very unlikely that the amplitude of the transmitted field be linear in ϑ .

On the other hand if, contrary to the immediate feeling, ϵ'_T/T were shown to be $\propto \vartheta$, one could still relax the mechanical specifications posed by (26) by increasing the distance between the two orthogonal resonant modes.

This can be done either by moving, at the expense of coupling, the input reference plane in the Tee where Γ_3 is forced to be real (its position is determined by the length of the line connecting the two end cavities), or by modifying the Tee in such a way as to increase $|\delta|$ (tip R+M towards -1). This can be done by inserting in the Tee an obstacle which acts as a mode filter. In Fig. 3 different structures are shown to illustrate how R+M is rotated.

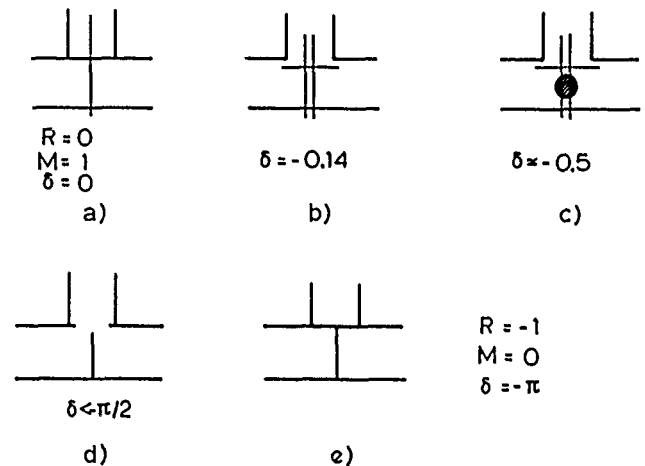


Fig. 3 - Possible realizations of a symmetrical E-plane Tee structure with indications of the rotation of the R+M vector. Cases a) and e) are obviously limit situations.

Cases a) and e) are theoretical, case b) is calculated in the appendix and cases b) and c) were experimentally measured with a network analyzer.

The use of a structure like c) or d) would guarantee enough separation between the two modes to fix the position of the stationary phase point even if ϵ'_T/T should be shown to be proportional to ϑ , however it is the author's feeling that only ϵ''_T/T is linear in ϑ and that the unmodified open E-plane Tee can be safely used in a microwave ring of a few wavelengths.

Conclusions

A new concept for the reduction of distributed phase shift in atomic beam frequency standards in the microwave region has been presented here with the proposal to replace the end shorts of the traditional Ramsey cavity with two equal ring cavities of a few wavelengths.

The problems raised by possible asymmetries in the rings were analyzed. It appears from the given analysis that the position of the stationary phase point can be predicted to fractions of a mm in a small ring.

Such a structure is capable of guaranteeing an average phase variation of less than 5×10^{-6} rad per mm displacement of the beam gravity center, which is equivalent to a frequency shift of less than 2×10^{-14} / mm for a primary Cesium standard with $Q = 10^8$.

Acknowledgements

The author would like to thank the many scientists who encouraged and helped him with discussions and constructive criticism. In particular L. Cutler of HP, R.F. Clark of NRC, G. Kramer of PTB, R. Drullinger, D.J. Glaze, and J. Shirley of NBS, G.P. Bava and I. Montrosset of the Polytechnic of Turin and E. Bava and A. Premoli of IEN.

Appendix

We calculate the scattering matrix of a symmetric lossless E-plane Tee from the equivalent circuit given in [10] and shown in Fig. 4 for reference planes at the corners of the Tee.

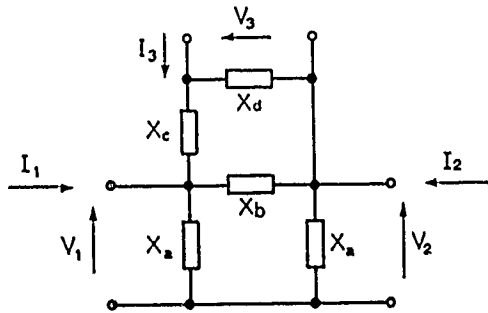


Fig. 4 - Equivalent circuit of the E-plane symmetric Tee for reference planes at the Tee corners.

From the latter, with the notation

$$Z = jX = j \begin{pmatrix} r & m & t \\ m & r & -t \\ t & -t & k \end{pmatrix} \quad (A1)$$

we find

$$\begin{cases} r = -m + X_a \\ m = f X_a \\ t = fg/X_c \\ k = f(2hX_d + g/X_a) \end{cases} \quad (A2)$$

where

$$\begin{cases} f = X_a / \left(2X_a + X_b \frac{X_c + X_d}{X_b + X_c + X_d} \right) \\ g = \frac{X_b X_c X_d}{X_b + X_c + X_d} \\ h = 1 - g/(X_b X_c) \end{cases} \quad (A3)$$

The scattering matrix is given by

$$\begin{cases} S = (jX-1)(1+jX)^{-1} = 1-2A-j2B \\ A = (1+X^2)^{-1}; \quad B = -XA \end{cases} \quad (A4)$$

The algebraic expressions of A and B as a function of X are given below for completeness. D indicates the determinant of $1+X^2$.

$$\begin{cases} D = (X_a^2 + 1)[(k^2 + 1)(m-r)^2 + (2t^2 + 1)^2 + k^2 + 4kt^2(m-r)] \\ A_{33}D = (X_a^2 + 1)[(m-r)^2 + 2t^2 + 1] \\ A_{13}D = (X_a^2 + 1)t(m-r-k) \\ A_{11} + A_{12} = 1/(X_a^2 + 1) \\ (A_{11} - A_{12})D = (X_a^2 + 1)(k^2 + 2t^2 + 1) \\ B_{33}D = -(X_a^2 + 1)[k(m-r)^2 + 2t^2(m-r) + 1] \\ B_{13}D = -(X_a^2 + 1)t[k(m-r) + 2t^2 + 1] \\ B_{11} + B_{12} = -X_a/(X_a^2 + 1) \\ (B_{11} - B_{12})D = (X_a^2 + 1)[(k^2 + 1)(m-r) + 2kt^2] \end{cases} \quad (A5)$$

With the numerical values which can be read from the plots given in [10] the following values can be calculated for the scattering parameters.

$$\begin{cases} |R| = 0.30; \quad \phi_R = -1 \\ |M| = 0.70; \quad \phi_M = -1.13 \\ |T| = 0.64; \quad \phi_T = -0.78 \\ |K| = 0.43; \quad \phi_K = -0.33 \end{cases} \quad (A6)$$

These are probably correct within a few percent. By moving the reference planes inward by l_K and l_T respectively we can make T and K real.

It must be $\beta l_K = 0.17$; $\beta l_T = 0.78 - 0.17$ or $l_K = 1.23$ mm and $l_T = 4.58$ mm, which puts the reference planes for the symmetric ports 1 mm apart.

It turns out that for these planes $\phi_R = -0.24$; $\phi_M = -0.11$ and $\delta = -0.14$.

References

- [1] S. Jarvis Jr.: "Molecular Beam Tube Frequency Biases due to Distributed Cavity Phase Variations" - NBS Tech. Note 660, Jan. 1975.
A. Bauch, T. Heindorff, R. Schroeder: "Measurement of the Frequency-Shift Due to Distributed Cavity Phase Difference in an Atomic Clock" - IEEE Trans. on Instr. and Meas., IM-34, 136-138 (1985).
- [2] A. De Marchi, G.P. Bava: "On cavity phase shift in Commercial Caesium-Beam Frequency Standards" - Metrologia 20, 33-36 (1984).

- [3] A. De Marchi, G.D. Rovera, A. Premoli: "Pulling by Neighbouring Transitions and its Effects on the Performance of Caesium Beam Frequency Standards" - Metrologia 20, 37-47 (1984).
- [4] A. De Marchi, R.E. Drullinger: "Long Term Behaviour of Cavity Phase Difference in NBS-6" - Metrologia, to be published.
- [5] D.J. Wineland: "The Cesium Beam Frequency Standard - Prospects for the Future" - Metrologia 13, 121-123 (1977).
- [6] R. Drullinger, J. Shirley, D. Glaze, L. Hollberg, A. De Marchi: "Progress Toward an Optically Pumped Cesium Frequency Standard" - Proc. 40th Freq. Contr. Symp. (1986).
- [7] L. Cutler: Private Communication.
- [8] E.O. Schulz-Du Bois: "Alternative Interpretation of Rotation Rate Sensing by Ring Laser" - IEEE Jour. of Quantum El., QE-2, 299-304 (1966).
- [9] See for example: D.M. Kerns, R.W. Beatty: "Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis" - Pergamon Press (1967).
- [10] N. Marcuvitz: "Waveguide Handbook" - McGraw Hill (1951).

A COMPARISON OF VARIOUS ALKALI GAS CELL ATOMIC FREQUENCY STANDARDS

J. C. Camparo and R. P. Frueholz

Chemistry and Physics Laboratory
The Aerospace Corporation
P. O. Box 92957
Los Angeles, CA 90009Abstract

In the present study a theoretical comparison among various alkali gas cell atomic frequency standards is undertaken, specifically: Rb^{85} , Rb^{87} , Cs^{133} , Cs^{135} and Cs^{137} . It is found that Rb^{87} exhibits the best potential shot noise limited performance of all the candidate alkalis (in a minimum volume TE₁₁₁ microwave cavity $\sigma_y(\tau) = 5.4 \times 10^{-15}/\sqrt{\tau}$), and that this is due to: i) a large ground state hyperfine splitting, and ii) a low value of nuclear spin. Additionally, the calculations indicate the importance of microwave cavity geometry on the stability one can attain with a gas cell atomic frequency standard.

I. Introduction

Though gas cell atomic frequency standards based on various alkalis have been discussed in the past [1]-[5], notably the Cs^{133} gas cell standard [3][4], none has ever come near to rivaling the popularity of the Rb^{87} gas cell standard. Obviously, there are several factors which contribute to this situation; for example, Rb^{87} has the second largest ground state hyperfine splitting of all the stable alkali isotopes (implying a relatively high atomic Q), and it has a low value of nuclear spin I so that optical pumping is relatively efficient in populating the $m_F = 0$ state. The primary reason for the supremacy of the Rb^{87} standard, though, lies with the fortuitous overlap of Rb^{85} and Rb^{87} optical absorption lines, which allows for the construction of a very simple and efficient hyperfine filter for optical pumping with alkali discharge lamps [6]. However, with the advent of narrowband lasers for optical pumping in atomic frequency standards [7] [8], especially single-mode diode lasers [9], the prerequisite of an efficient hyperfine filter in the design of a gas cell standard has been eliminated. This in turn has diminished the intrinsic attractiveness of a Rb^{87} gas cell standard. In particular, given the 35% greater ground state hyperfine splitting of Cs^{133} , one must seriously question the relative merits of a rubidium standard in the absence of a hyperfine filter requirement. In the present study it is therefore our desire to theoretically explore the shot noise limited performance of various laser pumped alkali gas cell standards, so that their intrinsic performance capabilities can, to a degree, be assessed.

Considering the fact that there are very many alkali isotopes, with various values of nuclear spin and various ground state hyperfine splittings, it is necessary to reduce this host of candidate gas cell standards by requiring the candidate alkali isotope to possess a few reasonable characteristics. Obviously, the first requirement is that the alkali isotope be either stable or long lived. Specifically, we must

require that the isotope have a half-life in excess of at least ten years. This is particularly important for space applications of frequency standards, where longevity is of prime importance. Furthermore, the alkali isotope must have half-integer nuclear spin, so that the standard can be based on a field insensitive 0-0 transition. Thus, considering a full range of alkali isotopes [10], one is left with the series of nine candidate alkalis collected in Table I. Additionally, in order to obtain efficient laser pumping one would require that the hyperfine resonances be well resolved optically, that is that the Doppler broadening be less than the ground state hyperfine splitting. Consequently, out of the host of original candidate alkalis, only the following five will be considered for further comparison [11]: Rb^{85} , Rb^{87} , Cs^{133} , Cs^{135} and Cs^{137} .

II. Overview of the Gas Cell Frequency Standard Model and Calculation

In previous publications [12] [13] we have discussed a non-empirical model of the gas cell atomic frequency standard. In brief, this model considers the relevant gas phase physics as occurring on two different scales. On what we term the "microscopic scale" the 0-0 hyperfine transition lineshape of an arbitrary alkali atom of half-integer nuclear spin is determined by the generalized Vanier theory of alkali atom hyperfine optical pumping [14] [15]. Among other parameters this theory considers the dependence of the hyperfine lineshape on optical pumping light intensity and microwave Rabi frequency. However, because the buffer gas pressure in a gas cell standard effectively freezes the alkali atoms in place on time scales of the order of a Rabi period [16], and because the alkali vapor is not necessarily optically thin; these two parameters, and hence the microscopic lineshape, vary from atom to atom within the vapor. Furthermore, as a result of diffusion to the resonance cell walls, where the atoms immediately depolarize on impact, there is a spatial distribution of hyperfine polarization $\langle \vec{I} \cdot \vec{S} \rangle$. In some sense this spatial distribution of $\langle \vec{I} \cdot \vec{S} \rangle$ can be imagined as being superimposed on the microscopic physics. Thus, there is also a "macroscopic scale" of physics in the problem which is related to the spatial variation of: (a) the optical pumping light intensity, (b) the microwave Rabi frequency, and (c) $\langle \vec{I} \cdot \vec{S} \rangle$ as a result of diffusion to the resonance cell walls.

In order to treat this macroscopic scale of physics in a reasonably lucid manner, the problem is reduced to one dimension, so that only the longitudinal variation of the optical pumping rate and microwave field strength is considered. This is reasonable because the microwave field can be made uniform in the transverse dimension by dielectrically loading the cavity [17], and because the laser intensity can easily be made uniform across the face of the resonance cell. The microwave Rabi frequency

distribution along the axial dimension is determined by the microwave cavity mode, assuming that the atomic resonance cell fills the microwave cavity. The axial variation of the optical pumping light intensity is determined by computing a "global" optical pumping parameter ξ in a self-consistent manner. In essence, this global optical pumping parameter determines the fractional population in the optically absorbing hyperfine multiplet, and thus the optical depth of the vapor as a result of optical pumping. Since the model assumes that the alkali atoms are effectively frozen in place in the resonance cell, the first order change in transmitted light intensity as a function of microwave Rabi frequency, for a uniform slice of vapor of thickness dz , only depends on the local values of the optical pumping light intensity and microwave Rabi frequency. In order to include the effect of axial diffusion this first order macroscopic solution is multiplied by the envelope function $f(z)$, which describes the axial distribution of hyperfine polarization in an optically thin vapor. (This procedure for treating diffusion is discussed more fully in Ref. 12.) When considering optical pumping with lamps, where the relative optical pumping rates are typically low, it is fair to approximate $f(z)$ by Minguizzi et al.'s first order diffusion mode [18]:

$$f(z) = \sin(\pi z/L), \quad (1)$$

where L is the length of the resonance cell. However, as previously discussed [13], when the optical pumping light source is a laser the envelope function must be generalized so as to be valid for arbitrary optical pumping rates:

$$f(z) = \frac{[1 - \exp(-\alpha z)] [\exp(\alpha L) - \exp(\alpha z)]}{[1 - \exp(-\alpha L/2)] [\exp(\alpha L) - \exp(\alpha L/2)]} \cdot (2)$$

In the present analysis we consider an alkali gas cell atomic frequency standard operating in its traditional configuration (i.e., cw optical pumping), except that the typical rf discharge lamp used for optical pumping is replaced by a single-mode laser tuned to the D_2 optical absorption resonance of the alkali $[n^2 P_{3/2} - n^2 S_{1/2}(F=I+1/2)]$. Furthermore, we assume that there is no filter cell, and that the resonance cell contains a pure isotopic vapor of the alkali under consideration. The calculations are performed in such a way that for a particular incident laser intensity we calculate the resonance cell temperature and peak microwave Rabi frequency that minimize the shot noise limited Allan variance as well as the minimum Allan variance value itself. Temperature enters the calculations through: 1) the temperature sensitivity of the alkali's diffusion coefficient, 2) the collision frequency of the alkali atoms (i.e., the relative speed of the atoms), and 3) the alkali vapor number density which determines both the spin exchange rate and the vapor's optical depth. Additionally, we consider a clock operating with a TE_{111} cylindrical microwave cavity of minimum volume:

$$L = c \sqrt{3}/2\nu_0 \quad (3a)$$

$$R = 1.841 \left[\left(\frac{2\pi\nu_0}{c} \right)^2 - \left(\frac{\pi}{L} \right)^2 \right]^{-1/2} \quad (3b)$$

where R is the cavity's radius, and ν_0 is the alkali's 0-0 hyperfine transition frequency. Thus, alkali gas cell standards with high hyperfine transition frequencies are modelled as operating with appropriately smaller microwave cavities and resonance cells. Other

general parameters used in the calculations are collected in Table II.

III. Results

The basic results of the calculations are presented in Fig. 1 and Table III. Clearly, Rb^{87} exhibits the best shot noise limited performance ($\sigma_y(\tau) \approx 5.4 \times 10^{-15}/\sqrt{\tau}$), and is roughly a factor of three better than any of the cesium isotopes. However, before saying too much about this result, one should note that the shot noise limited performance of the three cesium isotopes is a decreasing function of the isotope's hyperfine transition frequency. Given the equality of the nuclear spins of these isotopes, and the fact that the atomic line Q increases proportionate to ν_0 , this result is counter-intuitive; typically, we would expect $\sigma_y(\tau) \sim 1/\nu_0$ under the condition of equal isotope nuclear spin.

The resolution of this apparent paradox resides in the fact that these calculations assume a minimum volume microwave cavity, and hence a minimum volume resonance cell. As can be seen with the aid of Eqs. (3) the cell volume decreases like $1/\nu_0^3$. Thus, the cesium isotope results can be explained by postulating a negative correlation between microwave cavity size and the shot noise limited Allan variance. This hypothesis has been substantiated by running the cesium isotope calculations for the fictitious case of equal cavity radii and lengths: the resulting Allan deviations scaled like $1/\nu_0$ (i.e., Cs^{137} showed a 10% improvement in shot noise limited performance compared to Cs^{133}). Consequently, the results presented in Fig. 1 and Table III not only represent the intrinsic performance capabilities of the various alkali isotopes, but also geometrical effects associated with the microwave cavity's size.

How the cavity geometry affects gas cell frequency standard performance is at the present time not understood. One possible explanation is that by increasing the microwave cavity length, and hence the resonance cell length, one obtains good transmitted light signal amplitudes at relatively low alkali densities. Since low alkali densities imply low spin-exchange rates, and hence relatively higher atomic Q s and signal to noise ratios, longer resonance cells might be expected to exhibit improved clock performance. In the same vein, since larger resonance cell radii imply lower phenomenological diffusional relaxation rates, one again might expect some improvement in predicted clock stability with larger radii resonance cells. An alternative explanation, however, derives from the axial distribution of hyperfine polarization and microwave field strength in the resonance cell. Perhaps the influence of these distributions on clock signal amplitude is such that longer resonance cells yield relatively higher signal levels, and hence improved clock stability. In order to differentiate among these and other potential explanations for the influence of microwave cavity geometry on clock stability additional calculations need to be performed, preferably with a more rigorous 3-dimensional model of the gas cell atomic frequency standard.

To mitigate the influence of cavity geometry on the calculation of shot noise limited performance, we considered the fictitious case of a Cs^{133} gas cell frequency standard operating with a TE_{111} microwave cavity having the same radius and length as a Rb^{87} TE_{111} resonant cavity. Though the result of this calculation showed an improvement in clock stability compared to the Cs^{133} result given in table III, the best Allan deviation was nonetheless a factor of 1.6 larger than the corresponding Allan deviation of Rb^{87} . It would thus appear that the roughly factor of two difference in nuclear spins between Rb^{87} and Cs^{133} (i.e., the factor of two difference in the ground state degeneracies) more than compensates for the 35% difference in the alkali ground state hyperfine splittings.

It is also surprising that Rb^{85} should display such good performance compared to Rb^{87} considering that it has half the hyperfine resonance frequency, and a 50% greater ground state degeneracy resulting from its larger nuclear spin. In light of the preceding discussion of the Cs isotope Allan variances we attribute this relatively good performance to Rb^{85} 's ten times larger minimum microwave cavity volume. This again emphasizes the influence of the physics package geometry on the ultimate shot noise limited performance that can be attained with the gas cell standard.

IV. Summary

The present calculations indicate that a Rb^{87} gas cell standard shows the greatest potential for frequency stability, and in this regard nature has been uncommonly propitious. One should not, however, interpret this result as a superiority of the Rb^{87} standard in all regards. For example, if it is of primary importance to construct a miniature gas cell standard, then Cs^{133} might prove to be more advantageous given the fact that its minimum volume cavity occupies less than half the volume of a corresponding Rb^{87} cavity. Additionally, magnetic field sensitivities are less for Cs^{133} as a consequence of its greater hyperfine transition frequency. The only statement one should make regarding the present results is that of all the possible alkali gas cell standards one could consider, a Rb^{87} standard appears to yield the best attainable shot noise limited performance.

References

- [1]. W. E. Bell, A. Bloom and R. Williams, "A Microwave Frequency Standard Employing Optically Pumped Sodium Vapor," IRE Trans. Microwave Theory Tech. MTT-7, pp. 95-98, 1959.
- [2]. M. Arditi, "Frequency Control by Gas Cell Standards Fundamental Problems in the Light of Recent Experimental Results," Proc. 15th Annual Symp. on Freq. Control (Atlantic City, N.J., 31 May-2 June 1961) pp. 181-202.
- [3]. G. Rovera, A. DeMarchi and J. Vanier, "The Optically Pumped Passive Cesium Frequency Standard: Basic Theory and Experimental Results on Buffer Gas Frequency Shifts," IEEE Trans. Instrum. Meas. IM-25, pp. 203-210, 1976.
- [4]. G. Rovera, S. Leschiutta, G. Busca and F. Strumia, "Parameters Affecting the Stability of an Optically Pumped Cesium Frequency Standard," Proc. 32nd Annual Symp. on Freq. Control (Electronic Industries Assoc., Washington, D.C., 1978) pp. 466-468.
- [5]. C. Audoin and J. Vanier, "Atomic Frequency Standards and Clocks," J. Phys. E 9, pp. 697-720, 1976.
- [6]. J. Vanier, R. Kunski, P. Paulin, M. Tetu and N. Cyr, "On the Light Shift in Optical Pumping of Rubidium 87: The Techniques of 'Separated' and 'Integrated' Hyperfine Filtering," Can. J. Phys. 60, pp. 1396-1403, 1982.
- [7]. L. L. Lewis and M. Feldman, "Optical Pumping by Lasers in Atomic Frequency Standards," Proc. 35th Annual Freq. Control Symp. (Electronic Industries Assoc., Washington, D.C., 1981), pp. 612-624.
- [8]. G. Avila, E. deClercq, M. deLabachellerie and P. Cerez, "Microwave Ramsey Resonances From a Laser Diode Optically Pumped Cesium Beam Resonator," IEEE Trans. Instrum. Meas. IM-34, pp. 139-143, 1985.
- [9]. J. C. Camparo, "The Diode Laser in Atomic Physics," Contemp. Phys. 26, pp. 443-477, 1985.
- [10]. D. T. Goldman, in American Institute of Physics Handbook, edited by D. E. Gray (McGraw-Hill, New York, 1972), ch. 8b.
- [11]. It might be argued that since the ground state hyperfine sublevels of Na^{23} are resolvable, this alkali isotope should be considered for further comparison. However, since the nuclear spins of Na^{23} and Rb^{87} are equal, and since the ground state hyperfine splitting of Rb^{87} is roughly four times greater than Na^{23} , the only possible advantage to a sodium standard would be in the very high optical pumping rates that can be achieved with dye lasers at the sodium resonance wavelengths. Since it was shown in Ref. 13 that optimum shot noise performance is achieved at relatively low optically pumping rates, it does not appear that high laser intensities at the Na^{23} first resonance wavelengths will mitigate against Na^{23} disadvantages: non-completely resolved hyperfine transitions and a small ground state hyperfine splitting.
- [12]. J. C. Camparo and R. P. Frueholz, "A Nonempirical Model of the Gas-Cell Atomic Frequency Standard," J. Appl. Phys. 59, pp. 301-312, 1986.
- [13]. J. C. Camparo and R. P. Frueholz, "Fundamental Stability Limits for the Diode Laser Pumped Rubidium Atomic Frequency Standard," J. Appl. Phys. 59, pp. 3313-3317, 1986.

- [14]. J. C. Camparo and R. P. Frueholz, "Linewidths of the 0-0 Hyperfine Transition in Optically Pumped Alkali-Metal Vapors," *Phys. Rev. A* 31, pp. 1440-1448, 1985.
- [15]. J. C. Camparo and R. P. Frueholz, "Saturation of the 0-0 Hyperfine Transition Linewidth Enhancement Factor in Optically Pumped Alkali-Metal Vapors," *Phys. Rev. A* 32, pp. 1888-1889, 1985.
- [16]. R. P. Frueholz and J. C. Camparo, "Microwave Field Strength Measurement in a Rubidium Clock Cavity Via Adiabatic Rapid Passage," *J. Appl. Phys.* 57, pp. 704-708, 1985.
- [17]. H. E. Williams, T. M. Kwon and T. McClelland, "Compact Rectangular Cavity for Rubidium Vapor Cell Frequency Standards," *Proc. 37th Annual Symp. on Freq. Control* (IEEE Press, New York, 1983), pp. 12-17.
- [18]. P. Minguzzi, F. Strumia and P. Violino, "Temperature Effects in the Relaxation of Optically Oriented Alkali Vapours," *Nuovo Cimento* 46B, pp. 145-161, 1966.

Table 1. Some properties of the candidate alkali isotopes.

ALKALI	NUCLEAR SPIN I	Z ABUNDANCE OR HALF-LIFE	DOPPLER BROADENING (MHz)*	HYPERINE SPLITTING (MHz)
7_{Li}	3/2	92.58%	2340	804
23_{Na}	3/2	100%	1470	1772
39_{K}	3/2	93.10%	870	462
41_{K}	3/2	6.88%	850	254
85_{Rb}	5/2	72.15%	580	3036
87_{Rb}	3/2	27.85%	570	6835
133_{Cs}	7/2	100%	420	9193
135_{Cs}	7/2	3×10^6 yr	420	9724
137_{Cs}	7/2	30.0 yr	420	10,116

* Calculated Assuming a Vapor Temperature of 100°C

Table II. Parameters used in the calculation of the clock signal shot noise.

Parameter	Value
Laser Linewidth	50 MHz
Nitrogen Buffer Gas Pressure	10 torr
Photodetector Responsivity	0.5 amps/watt

Table III. Best Allan deviations for the various candidate alkalis as well as the corresponding laser intensity, cell temperature and peak microwave Rabi frequencies.

ALKALI	$\sigma_y^{\text{best}}(\tau)$	LASER INTENSITY $\mu\text{W}/\text{cm}^2$	TEMPERATURE $^{\circ}\text{C}$	PEAK RABI FREQUENCY Hz	MINIMUM CAVITY VOLUME cm^3
85_{Rb}	$5.5 \times 10^{-15}/\sqrt{\tau}$	90	67	97	338.0
87_{Rb}	$5.4 \times 10^{-15}/\sqrt{\tau}$	128	73	155	29.6
133_{Cs}	$1.3 \times 10^{-14}/\sqrt{\tau}$	371	79	380	12.2
135_{Cs}	$1.4 \times 10^{-14}/\sqrt{\tau}$	291	78	309	10.3
137_{Cs}	$1.4 \times 10^{-14}/\tau$	315	79	346	9.1

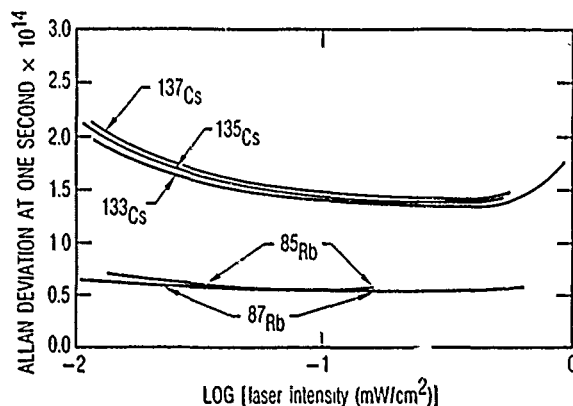


Figure 1: For the various alkali isotopes these curves show the best Allan deviations as a function of incident laser intensity.

Note that a Rb^{87} standard shows the best predicted performance. As discussed in the text the relation between the cesium isotope performances reflects the influence of cavity geometry on the Allan deviation, as well as intrinsic difference between the isotopes.

A RUBIDIUM-CRYSTAL OSCILLATOR (RbXO)

W.J. Riley and J.R. Vaccaro

EG&G Frequency Products
Salem, Massachusetts 01970ABSTRACT

The EG&G Rubidium-Crystal Oscillator (RbXO) is an oven-controlled Crystal Oscillator (OCXO) that is periodically syntonized by a rubidium (Rb) reference. The RbXO uses a digital tuning memory to hold the OCXO frequency control voltage while the Rb reference is off. This concept combines the best features of a rubidium frequency standard (fast warmup and low drift) with those of a crystal oscillator (low power and small size). The long-term stability of the Rb reference is transferred to a low-power crystal oscillator, forming a unit that is ideal for many tactical timing applications. The RbXO program was an 18-month R&D effort sponsored by the U.S. Army LABCOM. It included the design, manufacture, and test of eight RbXOs and three RbXO Demonstrators. An important aspect of the RbXO program was a 180-day Design Verification Test (DVT), during which four RbXOs were on-off cycled twenty times per day while being subjected to a -62 to +68°C temperature cycle. This test successfully verified that the design is capable of the desired 20-year life duration. Of particular interest is the low syntonization energy (2 W-hr at +25°C) and the absence of any significant drift in the syntonized frequency during many thousands of on-off cycles. The long-term stability of the RbXO is essentially identical to that of the rubidium frequency standard operating continuously.

INTRODUCTION

The RbXO is an oven-controlled crystal oscillator (OCXO) that runs continuously and is periodically syntonized by a rubidium (Rb) reference that can remain off most of the time. This concept combines the best features of a rubidium frequency standard (fast warmup and low drift) with those of a crystal oscillator (low power and small size).¹ The long-term stability of the Rb reference is transferred to a low-power crystal oscillator, forming a unit that is ideal for many tactical timing applications.

The RbXO is shown in Figure 1; its basic block diagram is shown in Figure 2.

The Rb reference portion of the RbXO is a miniature rubidium frequency standard modified to control an external crystal oscillator. The OCXO portion of the RbXO also includes a digital tuning memory to hold the frequency control voltage while the Rb reference is off. The crystal oscillator and tuning memory may be separated from the Rb reference for portable applications.

Detailed RbXO specifications and the actual results obtained are shown in Table 1; the main features are:

1. Usable with 5 or 10 MHz OCXOs
2. Size: ≤ 80 in.³ plus OCXO
3. Operating temperature: -55 to +68°C
4. RFS temperature stability: $\leq 4 \times 10^{-10}$
5. RFS temperature retrace: $\leq 5 \times 10^{-11}$
6. RFS drift: $\leq 1 \times 10^{-12}$ /day
7. RFS power at -55°C: ≤ 23 W
8. Interface power: ≤ 80 mW
9. Radiation hardened to tactical levels

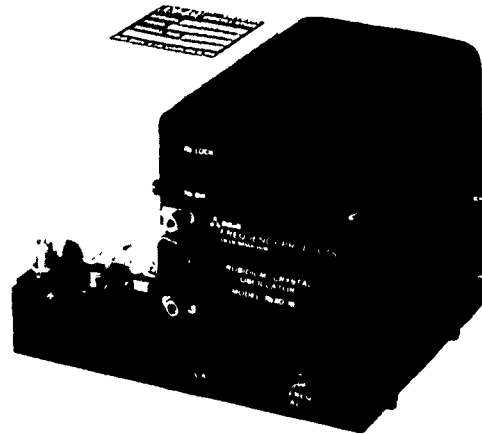


Figure 1. Rubidium Crystal Oscillator (RbXO)

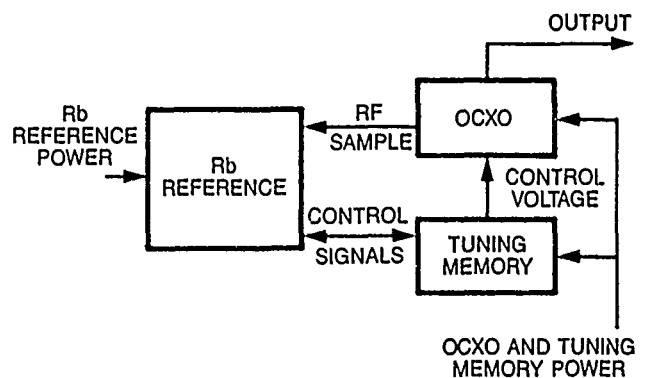


Figure 2. RbXO Basic Block Diagram

Implementation of the RbXO design was based on existing Rb reference and crystal oscillator hardware. The main technical concern was assuring that the Rb reference could endure thousands of on-off cycles. The RbXO program, therefore, included a Design Verification Test (DVT) that subjected four units to twenty on-off cycles and a -62 to +68°C temperature cycle each day for 180 days.

The major tasks of the RbXO development program included:

1. Design RbXO.
2. Build eight RbXOs.
3. Perform DVT on four RbXOs.
4. Design Demonstrator.
5. Build three Demonstrators.

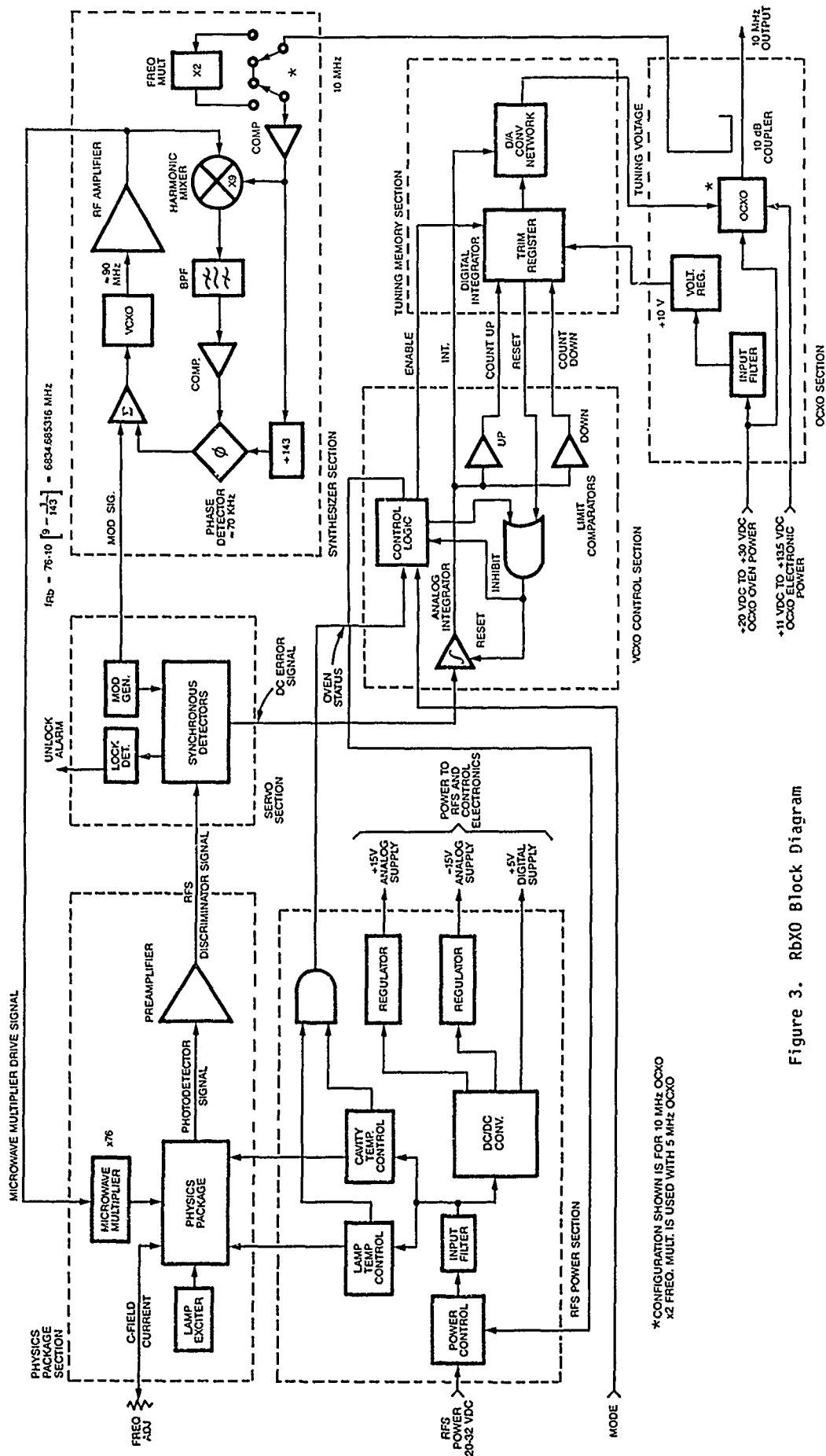


Figure 3. RbXO Block Diagram

These tasks were all successfully accomplished.² In particular, this program verified the efficacy of the RbXO concept and the ability of the hardware to endure severe on-off and temperature cycling while maintaining excellent frequency stability.

RbXO BLOCK DIAGRAM

A block diagram of the RbXO (based on the EG&G RFS-10 Rubidium Frequency Standard³) is shown in Figure 3. The major sections are:

1. The physics package that acts as a frequency discriminator to produce a signal which indicates the magnitude and sense of the difference in frequency between the applied rf excitation and the rubidium atomic resonance.
2. A servo amplifier that processes the discriminator signal to produce a dc error signal.
3. A VCXO control section that contains an analog integrator and limit comparators that adjust the VCXO tuning register, and overall control logic.
4. A tuning memory section with a digital register and D/A converter that tunes the VCXO.
5. The VCXO that produces the RbXO output.
6. A synthesizer section that converts the standard output frequency into an exact submultiple of the rubidium resonance.
7. A power section that provides supply voltages to the circuitry and has temperature controllers for the two physics package ovens.

The RbXO electronic diagram has a single frequency lock loop that contains a wide bandwidth PLL synthesizer as a subloop. A particularly simple rf chain is used that has no critical tuned circuits. The servo amplifier is a low-complexity cascade detector configuration. The power supply and temperature controller sections use efficient, high-frequency switching techniques.

The physics package section is identical to that of the RFS-10, except that the internal OCXO is omitted. The power and synthesizer sections are essentially unmodified. The normal rf output section is eliminated because an external OCXO is used. Minor changes were made in the servo section to interface with a new VCXO control section. The OCXO section contains a 10-MHz OCXO and a tuning memory section that provides the frequency control voltage for the OCXO. The OCXO control and tuning memory sections represent the most significant change from the conventional RFS. They replace the usual analog servo integrator with a hybrid analog/digital integrator that includes a digital register and D/A converter⁴. The digital register is an up/down counter controlled by the analog integrator and dual limit comparators.

During operation of the rubidium reference, the digital integrator provides a coarse control voltage. The analog integrator provides a fine control voltage that is summed to steer the OCXO to the exact center of the atomic resonance. The RbXO can be operated this way indefinitely and will be nearly indistinguishable from a normal RFS. The system is the usual frequency-lock loop arrangement with an external OCXO replacing the normal internal arrangement, and a hybrid analog/digital integrator replacing the normal analog version. OCXO drift results in a small dc error signal which causes the analog integrator to produce a ramp that is summed with the coarse digital integrator voltage to correct the OCXO.

When the analog integrator reaches its positive (or negative) limit, one of the comparators causes the digital

integrator to count down (or up). The analog integrator is also reset to its center to minimize disturbance of the OCXO.

Equilibrium is then reestablished and the process continues. There is only a slight transient frequency disturbance during the steps. The OCXO has a total tuning range of about 2×10^{-7} , and a 12-bit D/A converter is used; thus, the steps are about 5×10^{-11} . This tuning range requires annual coarse OCXO frequency adjustments to be made for an oscillator drifting about 3×10^{-10} /day. The OCXO can be expected to have a drift considerably lower than that when operated continuously for extended periods of time.

Changeover to standby operation is accomplished automatically or by removing the RFS power after syntonization. During standby operation, only the OCXO and its tuning memory are powered. The tuning register and D/A converter power requirements are very low since the register is CMOS and the D/A converter and summing network is a passive R/2R ladder network.

RUBIDIUM REFERENCE SECTION

The rubidium reference section of the EG&G RbXO is based on the Model RFS-10 Rubidium Frequency Standard. This miniature, militarized unit is well suited for the RbXO application. The RFS-10 is the smallest (30 in.³), lowest power (12W at +25°C) MIL-Spec Rb reference currently available. It also offers the lowest warmup energy (1.9 W-hr at +25°C to $\pm 2 \times 10^{-10}$) and steady-state power (18W at -55°C), important considerations for the RbXO application.

The main changes required in the RFS-10 for use in the RbXO include:

1. Elimination of the internal crystal oscillator.
2. Elimination of the rf output circuitry.
3. Provisions for 5 or 10 MHz excitation of the rf chain.
4. Modification of the servo amplifier for use with a digital integrator.
5. Addition of a control board to interface with the OCXO tuning memory.
6. Modification of the power board to allow automatic control of the Rb reference power.

The RbXO Rb reference includes the physics package, synthesizer, servo, control, and power sections as shown in Figure 3.

No changes were required in the basic RFS-10 physics package nor in the lamp exciter, microwave multiplier, photodetector, or preamplifier.

No changes were required in the synthesizer section itself. However, an optional X2 frequency multiplier was added and the normal rf output amplifier was eliminated from the RFS-10 output board, which became the RbXO divider board. The RFS-10 rf board is used without modification for the RbXO.

Minor changes were required on the RFS-10 servo and power boards for use in the RbXO.

The major change in the RFS-10 for the RbXO design was the addition of a fifth plug-in circuit board. This control board contains the logic to control the digital tuning memory and the Rb reference power.

TABLE 1. RBXO SPECIFICATION COMPLIANCE
(with Piezo 2810007 OCXO)

ITEM	REQUIREMENT	ACTUAL	ITEM	REQUIREMENT	ACTUAL
Output	10 MHz nominal	Same; 0.5V rms into 50 Ω Harmonics \leq -26 dBc Spurious \leq -100 dBc Phase Noise: f, Hz $\mathcal{L}(f)$, dBc/Hz 1 -90 10 -120 100 -140 1,000 -157 10,000 -160	Storage Temperature	-62 to +95°C	Same.
OCXO Compatibility	10 MHz TMXO, 10 MHz HP 10811 or equivalent, or any comparable stability 5 MHz OCXO	Same; Piezo Model 2810007 used (HP 10811 equivalent).	Temperature Stability	Rb reference: $\pm 4 \times 10^{-10}$	Same (2.5×10^{-10} typical).
OCXO Frequency Selection	Internal switch or jumper	Jumper		OCXO: not specified	$< 4.5 \times 10^{-9}$.
OCXO Frequency Control Time Constant	≥ 0.5 sec.	8.3 sec.	Warmup	Rb reference: ≤ 4 min. to $\pm 5 \times 10^{-10}$ at -55°C (with tradeoffs allowed)	8 min. at 26 Vdc (9 min. to $\pm 2 \times 10^{-10}$). Faster warmup with higher supply voltage or optional boost heaters as trade-off versus higher demand power and increased thermal stress.
Syntonzation Control	Automatic, on application of power to Rb reference, with automatic shut-off of Rb reference plus means to operate continuously in Rb mode.	Same; contact closure to ground keeps unit in Rb Mode. Automatic shutoff logic senses frequency stabilization.		OCXO: not specified	10 min. to $\pm 5 \times 10^{-9}$ at 25°C.
Syntonzation Error	$\pm 5 \times 10^{-11}$ max.	Same with Rb off. Infinite resolution with Rb on.	Rb Reference Retrace and Hysteresis	$\pm 5 \times 10^{-11}$	Same. ($\pm 1 \times 10^{-11}$ typical).
Monitoring	Indication when Rb reference is on and when OCXO is locked to it.	LEDs for Rb ON and Rb UNLOCK. Analog monitors for Rb light/signal and lamp and cavity ovens. Analog monitors for OCXO crystal oven and control voltage.	Aging	Rb reference: $\leq 1 \times 10^{-12}$ /day after 30 days	Same.
OCXO Mounting	OCXO and tuning memory separable for portable applications.	Same; OCXO and tuning memory approx. 27 in. ³		OCXO: not specified	$< 5 \times 10^{-10}$ /day after 24 hours.
Frequency Change Due to OCXO Separation	Less than $\pm 5 \times 10^{-10}$	Same ($\pm 2 \times 10^{-10}$ typical).	Power	Rb reference: ≤ 23 W at -55°C ≤ 110 W during warmup	20-32 Vdc 20W at -55°C; 33W during warm-up at 26 Vdc.
Rb Reference and Interface Size	≤ 80 in. ³	41 in. ³ (64 in. ³ with OCXO)		Tuning memory: ≤ 80 mW	Same. (70 mW typical).
Operating Temperature	-55 to +68°C	Same.		OCXO: not specified	2.5W typical at +25°C.
			Syntonzation Energy	Minimum	2W-hr at +25°C 4.5W-hr at -55°C
			Life	Accuracy of $\pm 3 \times 10^{-9}$ for 20 years.	Same; no wearout mechanisms.
			Radiation Hardening	Survivable to levels required of ground tactical systems	Same.
			TRFS Compatibility	Designed for commonality and compatibility with tactical Rubidium Frequency Standard (TRFS)	Same; physics package, block diagram, and most parts and subassemblies essentially identical to TRFS.

CRYSTAL OSCILLATOR SECTION

The crystal oscillator section of the RbX0 consists of an ovenized crystal oscillator and associated tuning memory.

The tuning memory adjusts the control voltage of the OCXO so that its output is locked to the Rb reference. The memory board maintains this control voltage after the Rb reference has been turned off.

The crystal oscillator section has three connectors: dc power, Rb reference interface, and rf output, and is separable from the Rb reference section of the overall RbXO.

A Piezo Model 2810007-1 Crystal Oscillator was selected by LABCOM as the crystal oscillator for the RbX0 development program.⁵ This unit is a high-stability, low-noise, ovenized 10 MHz oscillator that uses a SC-cut crystal. It is a readily available standard product having excellent performance as well as reasonable size and power.

The RbXO specifications require a design "capable of interfacing with either a 10 MHz Tactical Miniature Crystal Oscillator (TMXO), or any 10 MHz Hewlett-Packard 10811 equivalent OCXO or any comparable stability 5 MHz OCXO." The Piezo Model 2810007-1 is a HP 10811 equivalent.

The EG&G RbXO design is capable of interfacing with either 5 or 10 MHz oscillators by changing an internal

jumper wire, or of interfacing with the TMXO or other physically different oscillators by using a different tuning memory module. No change is required in the Rb reference section. Changes are required in only the tuning memory module to make the RBXO compatible with crystal oscillators having differing mechanical and electrical characteristics.

A TMXO was provided on loan from LABCOR to confirm its compatibility with the EG&G RbXO design. The small size (1.5 in.³) and low power (0.15 W at +25°C) of the TMXO makes it an ideal crystal oscillator for eventual use in a production RbXO. One non-DVT RbXO was delivered with a TMXO instead of a Piezo OCXO.

SYNCHRONIZATION LOGIC

A plot of a typical syntonization cycle is shown in Figure 4. In a normal RFS, the rubidium signal rises sufficiently to allow lock-up before the ovens are in control. In the RbXO, servo tracking is inhibited until a delay period of typically 90 seconds after oven control, so that the crystal oscillator frequency is not disturbed until the Rb reference has achieved an accuracy of better than 1×10^{-9} . The servo is then enabled and the crystal oscillator rapidly locks to and tracks the Rb reference. This process continues until the control logic senses stabilization (a frequency slope below a certain threshold) by the absence of servo counts for a certain time (typically 5 seconds). This control logic adapts to both oven warmup and frequency stabilization to provide minimum syntonization energy.

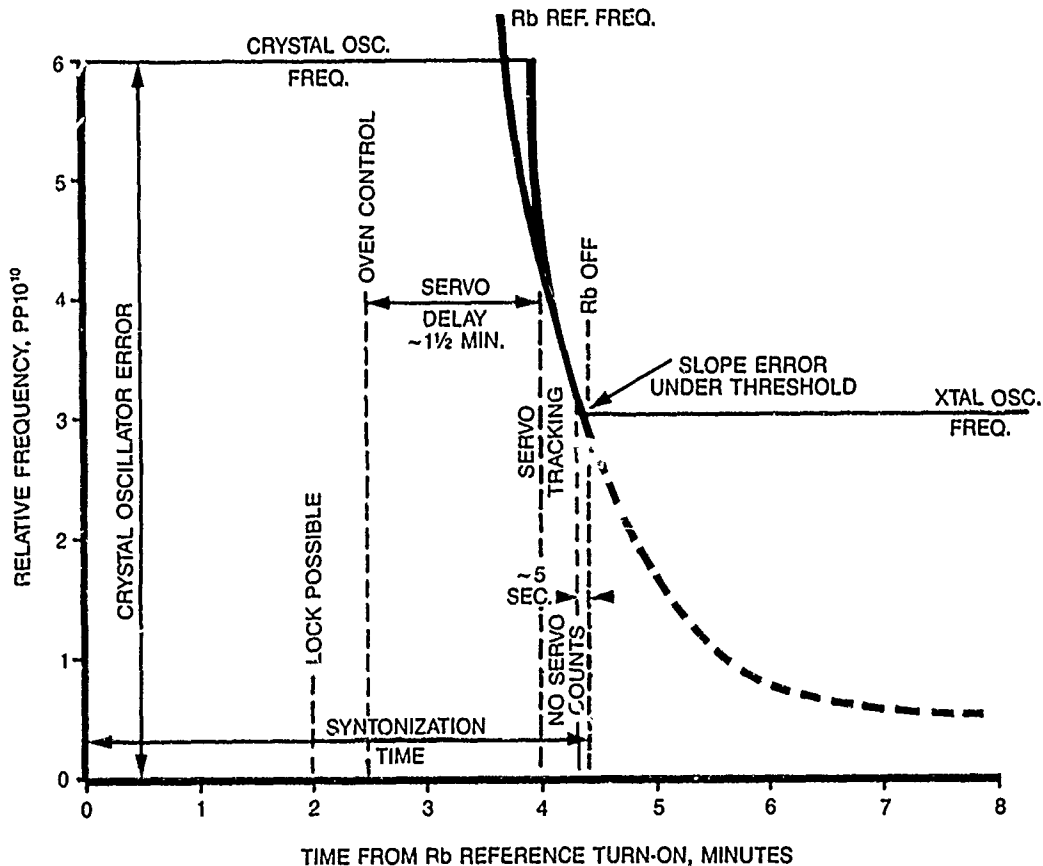


Figure 4. Typical RbX0 Syntonization Cycle

RbXO PACKAGING

Overall packaging of the RbXO is shown in the outline drawing of Figure 5. The package consists of two major sections, the main chassis which contains the Rb reference, and the OCXO with its associated tuning memory module. The OCXO mounts on the tuning memory module which, in turn, mounts to one side of the main chassis. Separate power connectors are used for the OCXO/tuning memory and the Rb reference. The tuning memory module has an output connector for the OCXO, and a connector that mates with an interface cable from the Rb reference.

The Rb reference is a modified EG&G RFS-10 rubidium frequency standard.

The RFS-10 has four main circuit boards that surround the physics package and plug into a motherboard at the front of the unit. The outer magnetic shield of the physics package has card guides that support the circuit boards. The outer shield and motherboard are supported by an L-shaped main structure. An outer cover completes the overall 2-3/4 inch square by 4 inch package.

The RbXO Rb reference has an additional circuit board added at the top to hold the control circuitry for the tuning memory. The heights of the motherboard and main structure are increased to about 3-1/4 inches to accommodate the additional board. The width of the main structure is also increased to about 5 inches to hold the crystal oscillator and tuning memory section.

The Piezo OCXO plugs into and is attached to the housing of the tuning memory, forming the crystal oscillator section (approximately 2 x 4 x 3 inches high). The entire crystal oscillator section is easily separated from the Rb reference section by removing four cap screws. This section has a separate power connector which allows its control interface to be disconnected from the Rb reference without interrupting the OCXO or its tuning memory.

The overall EG&G RbXO is 4.00 x 4.94 x 3.26 inches or 64.4 in.³ including the Piezo OCXO, and has a volume of only 41 in.³ without the OCXO, half the allowed size. The complete RbXO weighs 2.9 lbs.

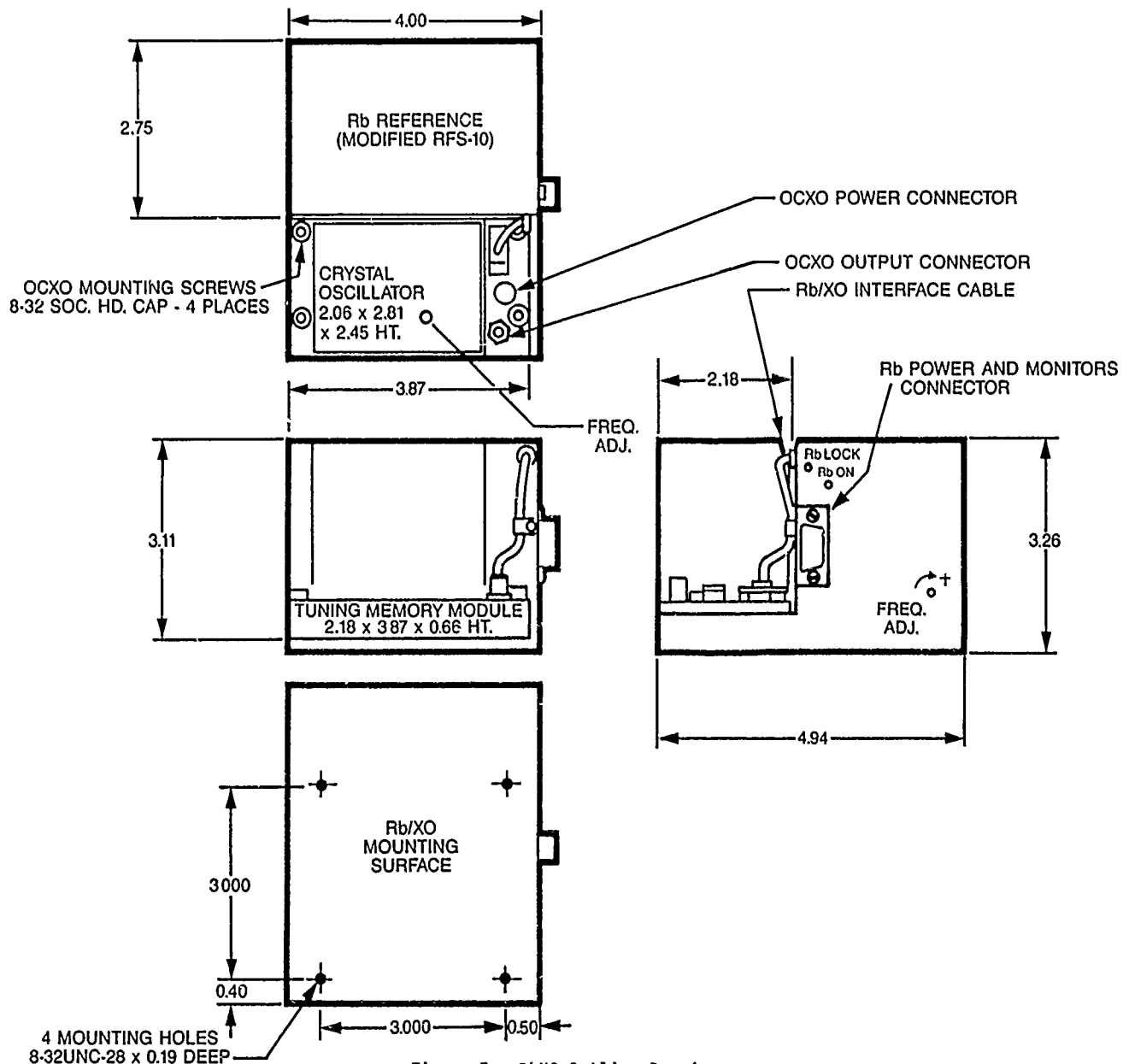


Figure 5. RbXO Outline Drawing

DEMONSTRATOR UNITS

The RbXO program effort included three deliverable RbXO demonstrator units. Each of these consists of a clock and timer to which an RbXO may be connected to demonstrate its function as a precision clock.

The digital clock displays time-of-day in hours, minutes, and seconds and has a phase-shiftable and synchronizable 1 pps output. The timer turns on the Rb reference continuously, or at an adjustable interval from 1 to 999 hours, as set by a thumbwheel switch.

The demonstrator units also contain a power supply and metering circuits and have space for mounting the RbXO. The power supply powers both the timer and RbXO from either the ac line or from an external battery. A block diagram of the RbXO demonstrator unit is shown in Figure 6.

A Time System Technology Model 6459 High Performance Digital Clock was selected as the digital clock for the demonstrators.⁶ This unit meets all the requirements for the digital clock and, as an off-the-shelf item, did not require any custom design effort.

The timer counts 1 pph pulses from the digital clock. It has logic to compare that count with the setting of a thumbwheel switch and thus controls the application of power to the Rb reference.

The power supply produces +28V from the ac line to run the Rb reference. The crystal oscillator oven is powered from either that source or an external +24 to +27V battery. The timer and crystal oscillator electronics are powered via a +12V regulator from either the +28V ac line supply or the external dc supply.

The status of the Rb reference and the crystal oscillator is monitored by a metering circuit. Additional circuits are included to provide 5 and 10 MHz outputs from the RbXO.

An opening on the timer panel provides space for mounting an RbXO. The RbXO can be seen and demonstrated when installed in the panel. The entire RbXO or its crystal oscillator section can be removed easily. Extender cables are also provided so that the RbXO can be operated outside the demonstrator.

The overall demonstrator is packaged in a 9-1/2-inch high bench rack cabinet, as shown in Figure 7.



Figure 7. RbXO Demonstrator

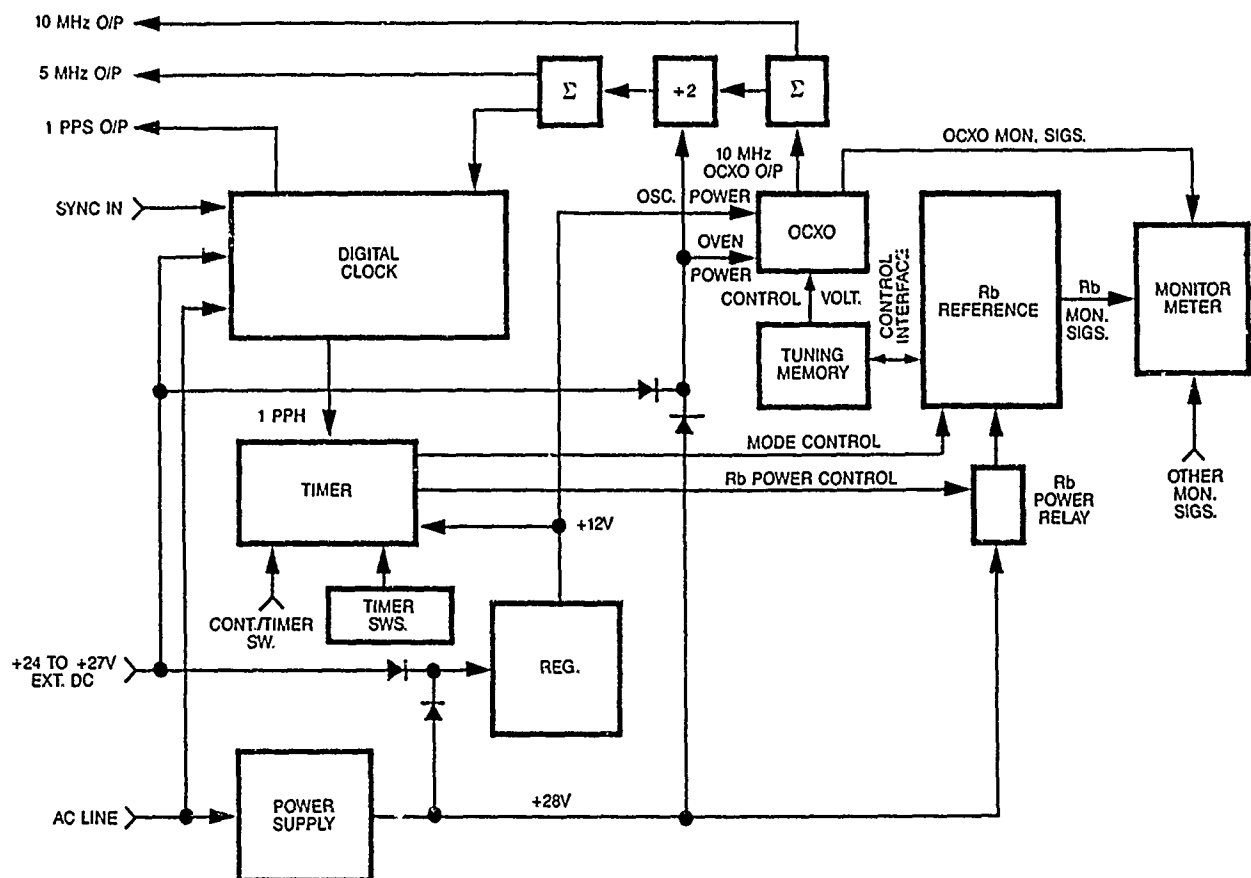


Figure 6. Demonstrator Block Diagram

RbXO ACCEPTANCE TESTS

Acceptance tests were conducted on all RbXO units. These tests included the normal Rb reference ATP items and additional items (such as syntonization energy) pertaining to the RbXO. These test results are shown in Table 2.

temperature of +68°C. Hourly $\tau = 100$ sec frequency measurements were made, from which daily averages were calculated and fitted to a log model. This data is shown in Figures 8 through 11. Daily monitor readings were also recorded.

Aging tests for the four DVT RbXOs were conducted in a temperature chamber set to provide a baseplate

Similar data were taken for the four non-DVT units.

TABLE 2. RbXO ACCEPTANCE TEST RESULTS

Parameter	Units		Spec	RbXO S/N							
				101	102	103	105	106	107	108	110 (TMXO)
Supply Voltage	V		----	28	24	27	27	27	27	27	26.6
Demand Power	W		≤110	36.1	29.1	33.2	33.7	34.0	34.8	36.6	33.5
Steady-State Power at -55°C	W		≤23	21.0	21.0	20.3	20.3	20.2	19.4	20.3	19.2
Syntonization Time	min. to ±2x10 ⁻¹⁰	at +68°C	----	3.2	3.8	3.4	3.3	3.0	2.5	4.6	2.7
		at +25°C	----	4.2	5.7	4.7	4.5	4.1	3.8	5.5	4.3
		at -55°C	≤4 to ±5 x 10 ⁻¹⁰ with tradeoffs	7.7	11.4	8.2	8.4	7.9	7.8	9.3	8.7
Syntonization Error wrt. SS value	pp10 ¹⁰	at +68°C	----	2.0	0.6	0.4	0.4	2.0	0.4	3.2	0.2
		at +25°C	----	3.0	0.4	3.0	1.0	1.0	2.4	0.5	0.4
		at -55°C	----	0.5	0.1	1.2	0.8	0.4	0.2	0.5	0.2
Frequency Difference after Syntonization	pp10 ¹¹		5	3	5	4	5	5	5	3	5
Syntonization Energy	W-Hr	at +68°C	----	0.9	1.0	0.8	0.9	0.8	0.6	1.2	0.7
		at +25°C	----	1.8	2.2	1.8	1.9	1.7	1.7	2.1	1.7
		at -55°C	----	4.0	5.0	4.0	3.8	3.8	3.8	4.5	4.1
Temperature Stability -55°C to +68°C	pp10 ¹⁰		±4	1.5	2.0	1.2	2.0	3.8	2.1	2.5	2.6
Voltage Sensitivity +25V to +30V	pp10 ¹¹		----	-0	-0	-0	2	-0	2	1	2
Magnetic Susceptibility	pp10 ¹¹ /Gauss		----	2.0	2.6	1.6	1.6	2.3	0.6	0.8	1.3
OCXO Separation Offset	pp10 ¹⁰		≤5	3	0.5	----	3.5	1.0	1.0	0.5	0.5
Trim Range	pp10 ⁹	Max.	----	+2.3	+1.9	+3.1	+1.0	+1.4	+1.7	+2.4	+2.9
		Min.	----	-4.8	-5.0	-2.3	-3.8	-3.5	-4.3	-2.1	-1.8
Drift	pp10 ¹³ /day		±10 after 30 days	+8.6	+9.6	-5.4	+8.2	+4.6	+23.6	+7.4	+5.9

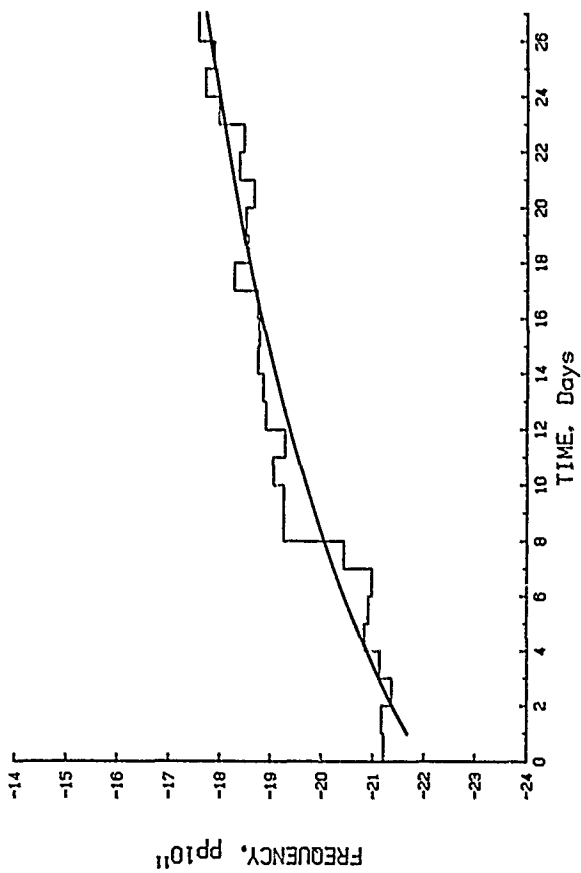


Figure 8. RbX0 S/N 101 Aging Plot

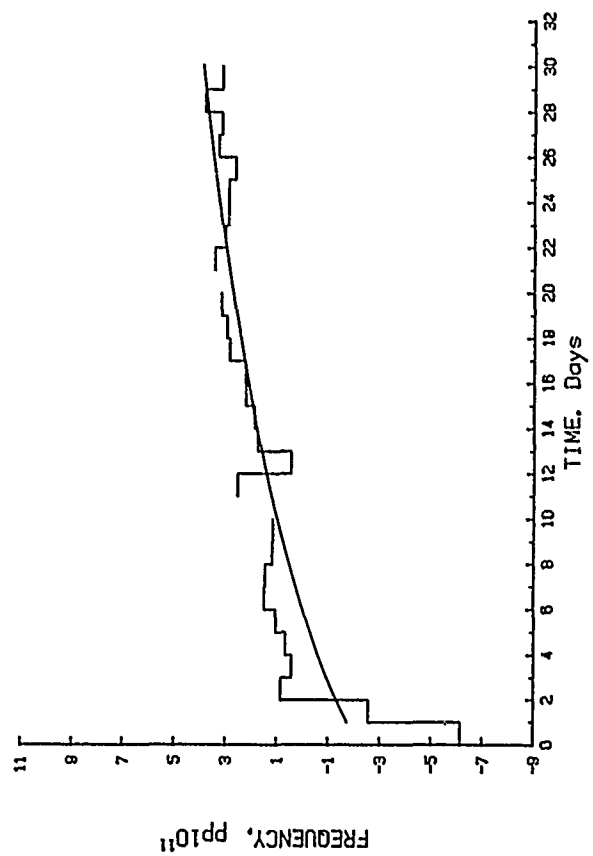


Figure 9. RbX0 S/N 102 Aging Plot.

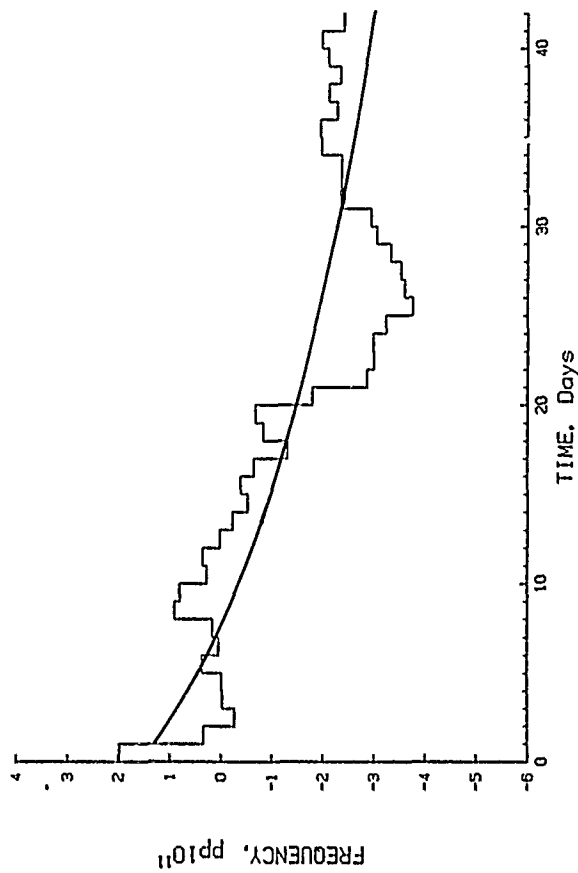


Figure 10. RbX0 S/N 103 Aging Plot

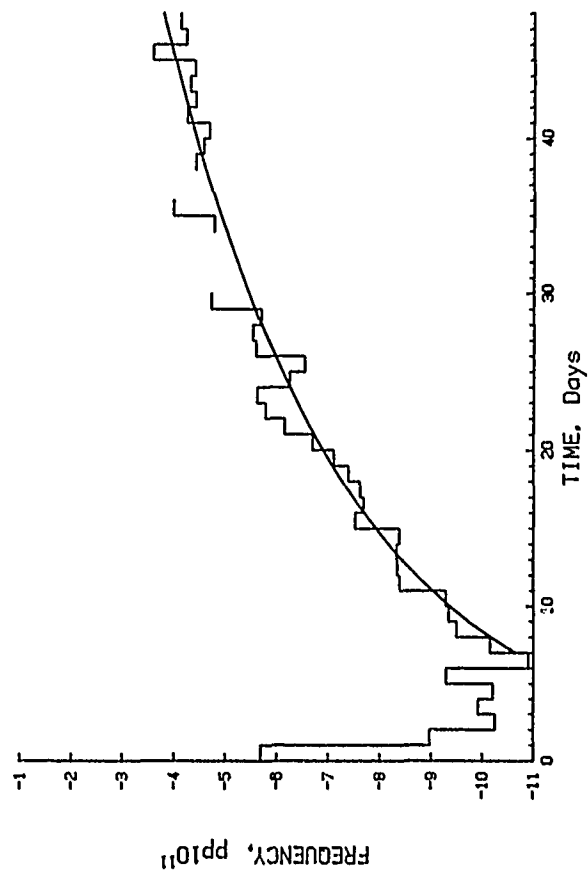


Figure 11. RbX0 S/N 108 Aging Plot

DESIGN VERIFICATION TEST

An important aspect of the RbXO development program was a 180-day Design Verification Test (DVT) during which four RbXOs were on-off cycled twenty times per day while being subjected to a -62 to +68°C temperature cycle. This test simulated the long-term usage of the RbXO, identified potential failure modes, and established RbXO reliability. The desired 20-year, maintenance-free life is possible because the Rb reference can withstand many thousands of on-off cycles and the attendant thermal cycling fatigue.

The DVT required a special test setup that included the following major items of test equipment:

1. A desktop computer to control the test system.
2. A programmable temperature chamber to hold the four RbXO units under test. This chamber was programmed to provide specified temperature cycles, and had auxiliary outputs to control power to the Rb references.
3. A frequency measuring system to measure the OCXO frequencies after each syntonization. The computer served as a data logger for this data.
4. A battery-powered backup system that ensured continuous operation of the four OCXOs. Interruptions of the Rb references and/or the temperature and power cycling because of a power failure was considered tolerable.
5. Other instrumentation to verify chamber temperature, monitor RbXO operation, and measure RbXO frequency versus temperature characteristics, Rb reference cell temperatures, and RbXO energy consumption.

The RbXO DVT extended over a nine-month period from late June, 1985, through mid March, 1986. Two RbXOs (S/N

102 and 103) ran flawlessly throughout their full 180-day tests. RbXO S/Ns 101 and 108 required minor repairs during tests, and S/N 108 experienced a failure that ended its test 10 days early. In addition, S/N 101 had an intermittency that caused several bad syntonizations. There were no OCXO problems and no Rb reference design deficiencies. All RbXO problems were workmanship defects associated with the developmental hardware. No unit showed signs of degraded performance during its 6-month DVT interval.

The test system had several partial outages, mainly associated with the temperature chamber and its control interface. The chamber cooling failed on two occasions for several days and its IEEE-488 bus interface hung up the measurement system for a day or so on three occasions. In addition, there were several other minor test and measurement system problems. None, however, had any significant effect on the overall test results.

A summary of the DVT log for each RbXO unit is shown in Table 3.

Plots of RbXO frequency during the DVT are shown in Figures 12 through 15 and are probably the most significant test results. They are plots of the average of the 20 daily frequency syntonizations and show the "RbXO drift", the trend in the retrace characteristic for the Rb references under on-off and temperature cycling. The plots include all days for which there was complete data. This is data never before available, and shows low drift for all units. This is the essence of the RbXO objective, to provide a stable, long-term frequency reference while consuming low energy.

TABLE 3. RbXO DVT SUMMARY

RbXO S/N	Start Day	End Day	Number of Days	Anomalies	Total On-Off Cycles	Total No. of Temp. Cycles	Drift $\mu\text{p1013/day}$	Total No. DVT Op. Hours
108	6/25/85 176	12/12/85 346	170	Intermittent connection re-soldered 10/23/85. Thermistor wiring failure ended DVT prematurely.	3250	160	+4.55	4080
103	8/12/85 224	2/10/86 406	182	Aborted test start 8/9/85 through 8/12/85. No anomalies during L/T period.	3547	175	-2.03	4440
102	8/26/85 238	2/24/86 420	182	No anomalies before or during DVT period.	3515	174	+1.66	4380
101	9/16/85 259	3/19/86 443	184	Aborted test starts 7/25/85 through 9/11/85. Intermittent connection re-soldered 3/6/86.	4335	209	+3.36	5304
Totals	----	----	----	One failure	14,647	718	----	18,204

Average Drift = +3.36 pp10⁻¹³/Day

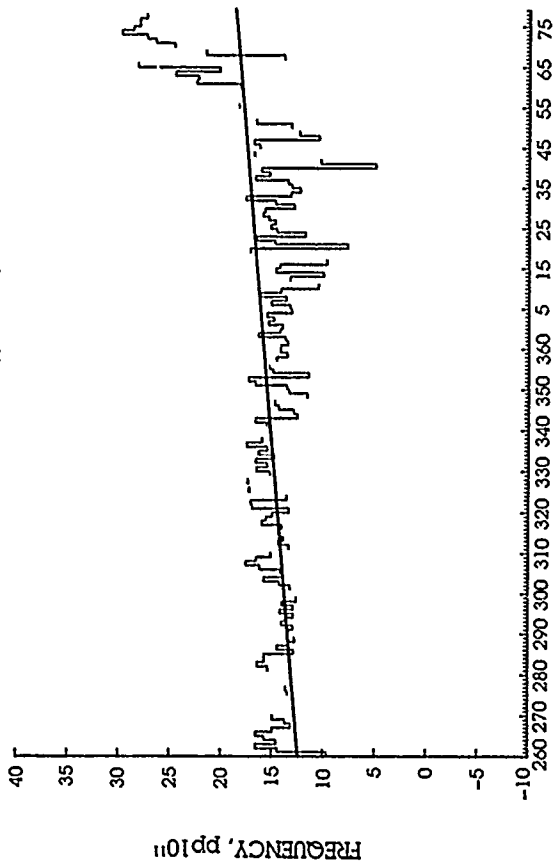


Figure 12. RbXO S/N 101 DVT Drift

Average Drift = +1.66 pp 10⁻¹³/Day

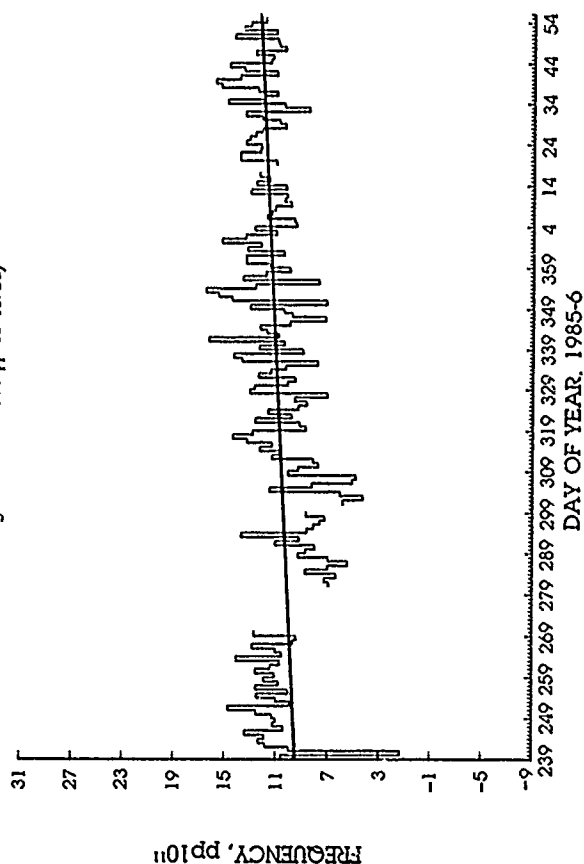


Figure 13. RbXO S/N 102 DVT Drift

Average Drift = -2.03 pp 10⁻¹³/Day

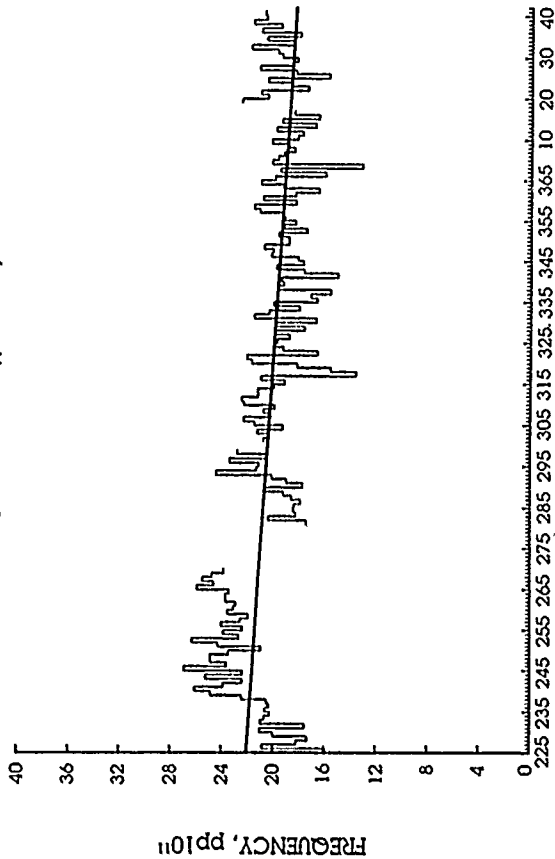


Figure 14. RbXO S/N 103 DVT Drift

Average Drift = +4.55 pp10⁻¹³/Day

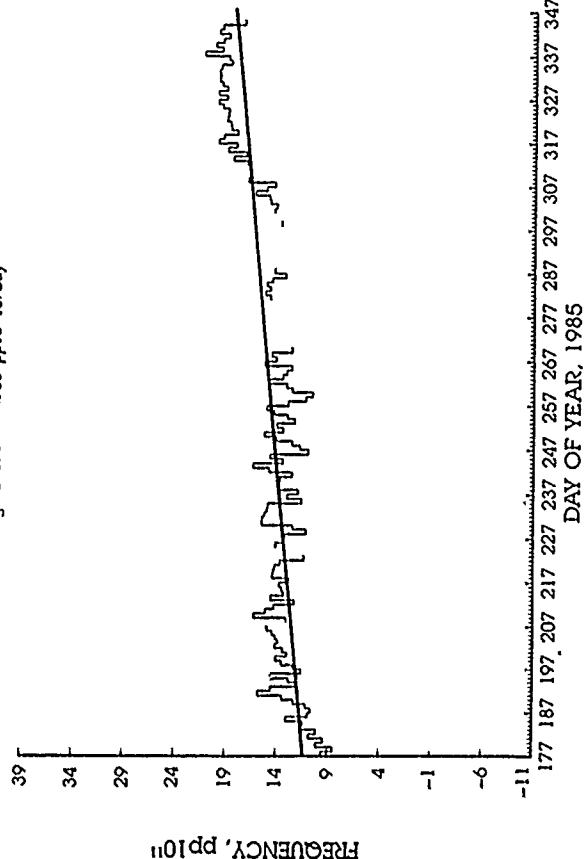


Figure 15. RbXO S/N 108 DVT Drift

Figures 16 and 17 show the maximum daily frequency deviations from nominal and from the first cycle for one (typical) RbXO.

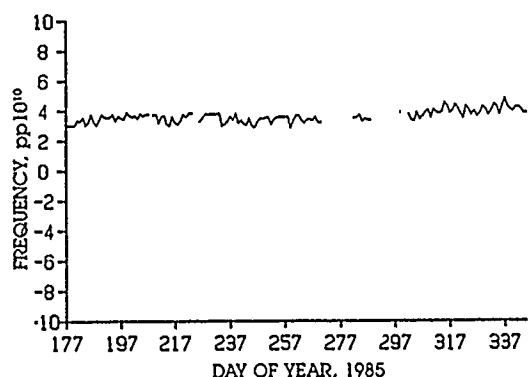


Figure 16. RbXO S/N 108 Maximum Daily Frequency Deviation from Nominal

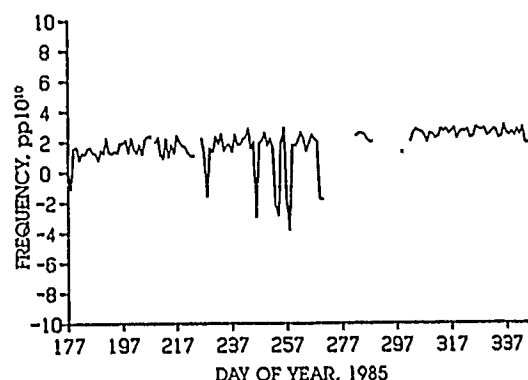


Figure 17. RbXO S/N 108 Maximum Daily Frequency Deviation from First Cycle

Figure 18 shows the average frequency versus temperature plot for this RbXO, the dependence of the syntonized frequency on the chamber air temperature due to the temperature characteristic of the Rb reference. It is the average of all frequency measurements corresponding to each of the 20 daily syntonizations, plotted against the nominal chamber air temperature, and is similar to that obtained under steady-state conditions. The hysteresis effect is due to the opposite slopes of the temperature ramps. There was no change in the RbXO temperature characteristics from the beginning to the end of the DVT (except for a slight displacement due to drift).

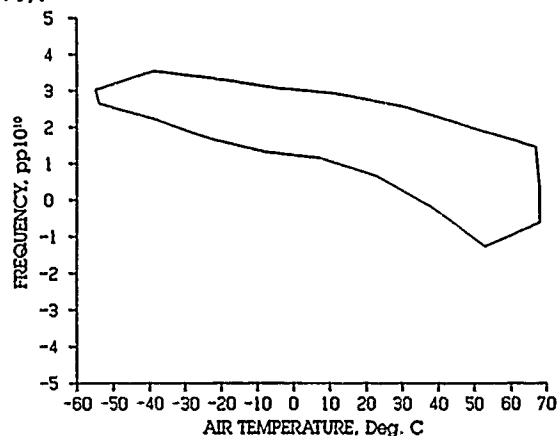


Figure 18. RbXO S/N 108 Average Frequency versus Temperature

Figure 19 is a plot of clock error for this RbXO. The clock error plot is a record of daily RbXO clock data, derived from integrating quasi-continuous $\tau \approx 10$ second scanned frequency measurements.

The clock error data was generally reset to zero when the measuring system was restarted after a major interruption. An alternative to this is a gap and displacement of the curve, such as seen at about day 220.

The most significant aspect of the clock error plots is their slope, which indicates the average RbXO frequency offset, including the uncorrected OCXO between syntonizations. For this unit the slope is about $+15 \mu$ sec/day, corresponding to an average frequency offset of about $+1.7 \times 10^{-10}$, very similar to that of the Rb reference.

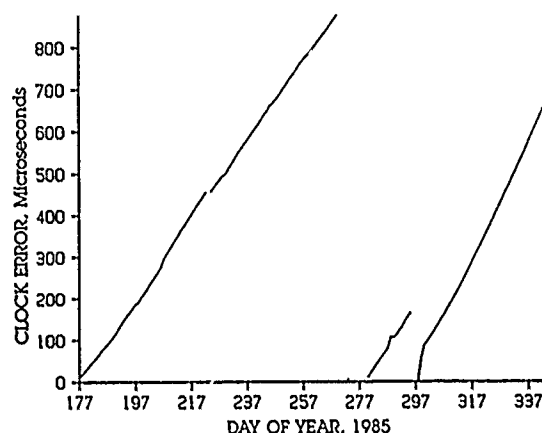


Figure 19. RbXO S/N 108 Clock Error

In addition to the frequency measurements made to determine RbXO clock error and syntonization accuracy, monitor readings were made to verify proper test system and RbXO operation throughout the test.

The internal temperature of the Rb references decayed essentially all the way to the air temperature between turn-ons, even at the -55°C end. Thus the one-hour syntonization cycle was adequate to fully exercise the RbXOs under test.

The DVT simulated, by means of severe temperature and on-off cycling, the 20-year service life expected for a RbXO operating in a tactical military environment. As such, the results were highly successful. Three units withstood the full test duration, and the fourth 94% of it. The latter failure testifies to the rigors of thousands of thermal fatigue cycles and the importance of small constructional details; in this case, an already recognized need for stress relief in a bead thermistor lead.

The most important DVT result was that a properly designed small rubidium reference can provide essentially the same long-term stability and reliability under conditions of severe temperature and on-off cycling as a unit operated continuously. This is new data, and it is encouraging for the expected large-scale usage of such devices in a tactical military environment.

SYNTONIZATION TIME AND ENERGY

Another important aspect of the RbXO performance is the time and energy required for the Rb reference to automatically syntonize the crystal oscillator (see Table 2). The syntonization energy is particularly important in an application where the RbXO must operate for many days on standby battery power.

The syntonization time is determined by the warmup and stabilization of the Rb reference and the criteria used by the control logic to automatically end the syntonization process.

The RbXO Rb reference requires only 5 minutes to warm up and stabilize to within $\pm 2 \times 10^{-10}$ of final frequency at +25°C ambient temperature and +26 Vdc supply voltage. Furthermore, because of efficient switching power circuits and low thermal mass, it requires only about 33W during warmup. The efficiency is essentially independent of supply voltage, which can be raised to decrease the syntonization time as shown in Table 4. Faster warmup is possible with boost heaters but this further increases demand power, thermal fatigue stress, and complexity and does not proportionally reduce the total syntonization time because of the need for internal thermal stabilization.

The EG&G RbXO uses adaptive control logic to determine the end of its syntonization process. This logic senses Rb reference oven stabilization and the settling of the OCXO control voltage to turn off the Rb reference as soon as possible at any ambient temperature. Frequency stabilization is estimated by observing the time between control voltage counts.

TABLE 4. SYNTONIZATION CHARACTERISTICS
VERSUS SUPPLY VOLTAGE (RbXO S/N 102) AT +25°C

Supply Voltage (Vdc)	Syntonization Time (Min)	Demand Power (W)	Syntonization Energy (W-hr)
32	3.75	45.1	1.8
30	4.15	41.4	1.9
28	4.35	37.0	1.9
26	4.75	33.2	2.0
24	5.08	29.5	2.0
22	6.35	25.1	2.2

CONCLUSIONS

The EG&G RbXO program accomplished all its objectives. In particular, it proved the viability of the RbXO concept by establishing that the long-term stability of a small rubidium reference is not degraded by severe on-off and temperature cycling.

The RbXO is a new class of frequency source and, as such, offers new possibilities to the systems designer. Some of the advantageous ways an RbXO can be used are shown in Table 5. It is likely that the RbXO will play an important role in future time and frequency systems.

TABLE 5. RbXO APPLICATIONS

Configuration	Advantage(s)
RbXO with attached OCXO; Rb reference off most of the time.	Low average power. Low drift. High reliability (Rb reference life extended in standby mode, and not essential for system operation).
RbXO with separate OCXO.	Small size, light weight, and low power for portable applications. Single Rb reference can serve many OCXOs (lower cost).
RbXO with separate Rb reference.	Rb reference maintenance and calibration without timing interruption.
Rb reference on most of the time.	Normal rubidium frequency standard performance. Can operate with OCXO only beyond environmental limits of Rb reference.

ACKNOWLEDGMENT

This work was supported by the U.S. Army Electronics Technology and Devices Laboratory (LABCOM).

REFERENCES

1. John R. Vio and Vincent J. Rosati, "THE RUBIDIUM-CRYSTAL OSCILLATOR HYBRID DEVELOPMENT PROGRAM," Proc. of the Sixteenth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, November 27-29, 1984.
2. Final Report, Rubidium-Crystal Oscillator Hybrid Development Program (RbXO), Contract No. DAAK20-84-C-0410, EG&G Frequency Products, Salem, MA 01970.
3. Instruction Manual, Model RFS-10 Rubidium Frequency Standard, EG&G Frequency Products, Salem, MA 01970.
4. T.K. Gregory, "New Cesium Beam Frequency Standard For Flight and Ground Applications," 31st Annual Symposium on Frequency Control (1977), pp 551-554.
5. Data Sheet, Model No. 2810007 Series High Stability DRAT Crystal Oscillator, Piezo Systems, Carlisle, PA 17013.
6. Data Sheet, Model 6459 High Performance Digital Clock, Time System Technology, Inc., Lake Grove, NY 11755.

A DISCIPLINED RUBIDIUM OSCILLATOR

A. MacIntyre and S. R. Stein
Ball Efratom Division
Broomfield, Colorado 80020

Abstract

A disciplined oscillator has been developed to combine the short-term frequency stability of a rubidium atomic oscillator with the long-term timekeeping ability of the Global Positioning System (GPS). A digital control loop is used to reduce the temperature sensitivity of the rubidium frequency standard, to provide frequency and time referenced to UTC and to make possible improved time prediction during periods when the GPS time reference is not available. Performance data for Efratom's Model RGR Disciplined Rubidium Oscillator are presented. The design approach is described in sufficient detail so that it may be applied in other situations.

I. Introduction

We use the term disciplined oscillator to describe a system consisting of a tunable local oscillator and an external frequency or time reference. The system makes use of the disseminated information to determine the performance characteristics of the local oscillator and to control its output thereby providing performance improvements compared to the use of either component separately. We have chosen the Global Positioning System as the reference because it is available worldwide and provides UTC with an accuracy of 0.1 microsecond. It is capable of supporting completely automatic and autonomous operation. The rubidium atomic oscillator is an excellent match to the GPS time reference. Its short-term frequency stability is comparable to high performance cesium beam frequency standards but its long-term frequency drift and frequency accuracy are quite limited. However, these are precisely the performance elements which may be improved by utilizing the information available from the reference. On the other hand, rubidium atomic oscillators reach their best frequency stability after many hours compared to 10 to 100 seconds for the best quartz oscillators. Thus a control loop for disciplining a rubidium oscillator will use a much longer averaging time than is possible with a quartz crystal local oscillator. The result is improved performance today when GPS satellites are visible only a portion of each day and superior immunity to any degradation of GPS transmissions which may result from future implementation of selective availability.

The selection of the local oscillator and reference led immediately to the choice of a digital control loop rather than the use of analog techniques. The GPS data is provided by the receiver in digital form and the long time constants involved in the control of a rubidium frequency standard would make subsequent analog signal processing exceptionally inconvenient. Digital signal processing makes it convenient to

change the loop characteristics in order to deal with the initial turn on and other transient conditions.

The form of the control loop is determined primarily by the choice of a rubidium atomic oscillator as the controllable element. Rubidium oscillators have appreciable temperature coefficient and long-term aging. Since these characteristics are deterministic, the use of feedforward control to supplement feedback provides improved performance compared to a pure feedback control loop. After feedforward compensation of the deterministic frequency changes, the residuals are stochastic and the optimum control equation may be calculated. The Auto Regressive Integrated Moving Average (ARIMA) technique was utilized to derive the parameters of the optimum control equation as a function of the performance of the rubidium oscillator. The ARIMA approach was selected for its compatibility with the sampled nature of the GPS time data, the availability of ARIMA noise models for atomic oscillators, and the availability of information needed to solve this problem.

Section II describes the Model RGR Disciplined Rubidium Oscillator. Details are given of the hardware, the control loop implementation, and the resulting performance data. Section III describes the calculation of the optimum control equation and some modelling results which demonstrate the behavior of the control loop as a function of the control loop parameters.

II. Equipment Design and Performance

The disciplined oscillator (Efratom Model RGR) was designed using a modular approach.

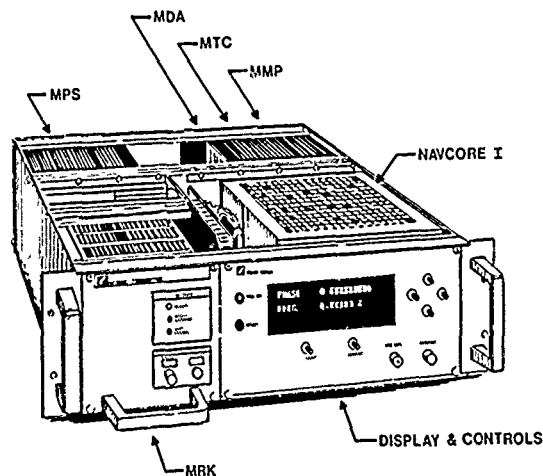


Figure 1.
Model RGR Rb-GPS Disciplined Oscillator
(Prototype)

A rubidium frequency standard, microprocessor, digital to analog converter, divider/counter and power supply are plugged into a standard 5 $\frac{1}{2}$ " high by 19" instrument rack containing a Navcore I GPS receiver manufactured by Collins Division of Rockwell International Corporation. Figure 1 is a line drawing of the complete system excluding the antenna and optional pre-amp, and Figure 2 shows the interconnection of the various modules.

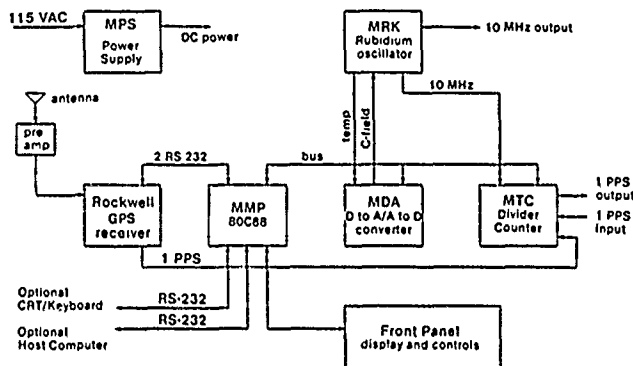


Figure 2.

Rb-GPS disciplined oscillator block diagram.

A variety of user interfaces are available. The front panel provides limited control and display capability and RS-232 ports are provided for optional CRT, keyboard and host computer interfacing.

All output signals are provided from the rubidium oscillator in normal operation. The 10 MHz signal is divided and phase shifted to provide the output 1 pps tick. The divider/counter compares the phase of this output signal with the free running 1 pps from the GPS receiver and transmits the raw time difference to the microprocessor (MMP). After the raw data is corrected using information provided from the GPS receiver, a new frequency is computed for the rubidium oscillator. Important performance characteristics of the modules are summarized below:

Modular Rubidium Oscillator. The MRK contains an Efratom Model FRK-H 10 MHz rubidium frequency standard. Its frequency may be varied

over a 2×10^{-9} range by adjusting the C-field using a 0 to 5 V DC input. Several BITE outputs are available for monitoring purposes and a temperature transducer is mounted in the heatsink to provide the necessary information for feedforward correction.

Navcore I GPS Receiver. This receiver provides the complete position, location and time reference function. The timing version locates the position of a fixed installation with 50 m accuracy and provides a reference to UTC with 100 ns accuracy. It has separate control and data buses which are managed internally by the modular microprocessor. The receiver has a single channel which is multiplexed to four satellites each second.

Microprocessor. The MMP module contains a microprocessor, memory, and four RS-232 interfaces. CMOS technology is used to conserve power and permit compact packaging. The memory consists of battery backed RAM for program development and EPROM for program storage. The

processor's burden comes mainly from supporting the RS-232 ports and the large number of floating point calculations needed for the statistical analyses. It operates at a speed of 5 MHz.

Converter. The MDA module contains two D to A converters and one A to D converter. One 12-bit DAC is used to control the rubidium frequency. This yields a resolution of

approximately 5×10^{-13} per bit. The second DAC is for future expansion. The analog to digital converter is multiplexed with eight 8-bit inputs. One is used for the temperature transducers and the others to monitor the BITE outputs of the rubidium oscillator.

The Divider/Counter. The MTC divides the 10 MHz from the rubidium oscillator and does phase shifting under control of the microprocessor. It performs time interval measurements to 12.5 ns resolution using a delay line method. An input multiplexer allows selection of either the GPS receiver 1 pps or user input.

The control loop operation is conventional in concept but some simple techniques have been found which reduce the acquisition time of the feedback loop and rapidly synchronize and syntonize the rubidium frequency standard. The first operation performed after the system is turned on is position location. The time required to complete this step is variable because of the limited number of GPS satellites presently available, but any delays at this step will be eliminated when the GPS system becomes fully operational. In any event, the position location operation is bypassed once a good position is determined and this operation is not repeated unless the receiver is moved. The position data used by the receiver is fixed during normal operation rather than updated with each satellite measurement in order to reduce the position location noise which shows up in the time information. The second operation is the calibration of the rubidium oscillator voltage tuning sensitivity. The frequency variation with control voltage is measured in a 200-second procedure which need not be repeated unless the rubidium oscillator is replaced or repaired.

The microprocessor begins closed loop operation by setting the rubidium oscillator to nominal frequency and initiating a sequence of frequency and phase measurements. Analysis of the timing data indicates that for sampling intervals from 1 second to 800 seconds white phase noise dominates. Consequently, optimum estimates of phase and frequency in this measurement region are obtained from the intercept and slope respectively of the straight line obtained from a linear least squares fit to the data. In the next section, it will be shown that a second order phaselock loop with appropriate parameters provides optimum control of the stochastic disturbances. It is also well known that a second order PLL will always pull in regardless of the initial frequency error, but the pull in time may be undesirably long. We have utilized two techniques to overcome this problem.

Any time error exceeding 100 ns during the initial lockup is removed by phase shifting rather than frequency tuning. The maximum start

up error of 0.5 s can be removed by this technique in 300 s. The frequency of the rubidium oscillator is intentionally offset from nominal at the end of the phase shifting sequence to remove the residual time error due to the 100 ns resolution of the phase shifter. Starting from this point, the microprocessor sequences through binarily increasing measurement intervals until the maximum 800 s measurement interval is reached. Standard second order PLL operation is not implemented at this time. Rather, the frequency corrections and phase corrections are maintained separately. In this way the integrator of the phaselock loop is not determined by the control necessary to remove the phase errors over the very short time intervals and a frequency estimate of ever increasing precision is obtained. After a two hour period a frequency calibration of a few parts in 10^{12} is obtained and second order loop operation commences. A single 800 second measurement each day suffices to maintain the closed loop system with a frequency accuracy of 5×10^{-12} and within a time error of 200 ns with respect to UTC.

The controller corrects for temperature induced frequency changes in addition to the feedback correction of the noise fluctuations. The temperature data is obtained from the transducer in the rubidium oscillator heatsink and the temperature coefficient is stored in nonvolatile memory. Although the temperature coefficient is thereby reduced by nearly an order of magnitude, temperature fluctuations remain the dominant perturbation for sampling times on the order of one-half day. The residual periodic frequency fluctuations disturb the control loop minimally when the attack time of the loop is chosen to be 24 hours. Figure 3 shows the time error in closed loop operation during approximately 10 days following initial turn-on. A 400 ns

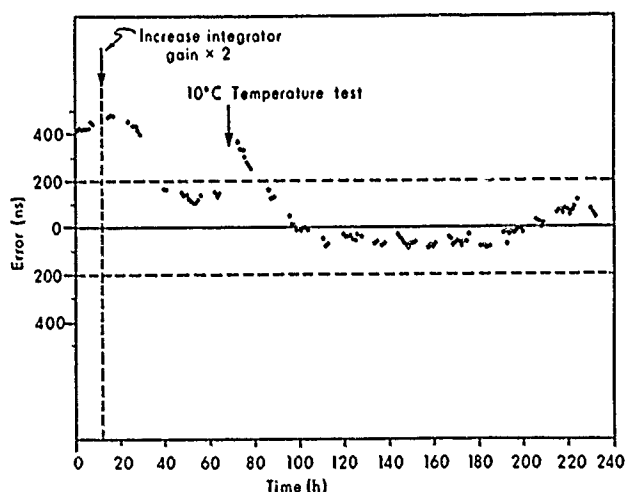


Figure 3.

Closed loop time error for ten days following turn-on.

initial time error was eliminated by increasing the integrator gain of the second order loop. Thereafter a maximum time error of approximately 300 ns was experienced as a result of a very rapid 10°C temperature change of the rubidium frequency standard. During the last five days of normal operation in a typical laboratory environment, the residual time errors never

exceeded approximately 100 ns. The data plotted in Figure 3 is the time error measured by the disciplined oscillator itself since an external reference was not available for this experiment. Figure 4 shows the measured frequency errors during the six days following turn-on of a

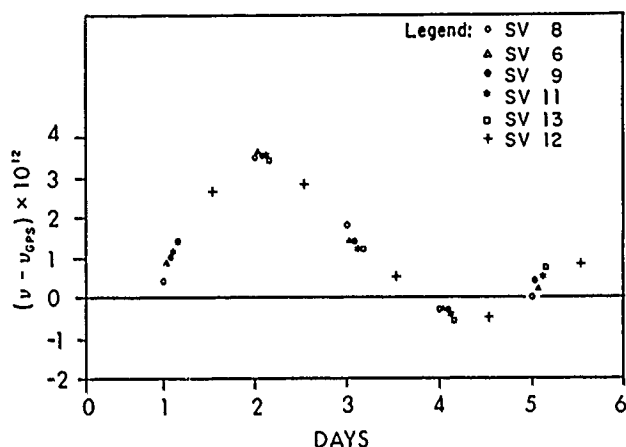


Figure 4.

Closed loop frequency error for six days following turn-on.

second experiment. The maximum frequency deviation observed during the turn-on transient was less than 4×10^{-12} . The frequency measurements were performed by an independent GPS receiver of the NBS design.

III. Control Loop Theory

The ARIMA methodology was selected to derive the control equation for the phaselock loop because of its simplicity and its prior application to the clock modeling problem. The general ARIMA approach is based on the premise that auto-regression of the stochastic quantity of interest may be equated to a moving average of current and prior random shocks¹. The model is applied to samples taken at equal intervals and may be written in the form of Equation 1 where $x(t)$ is the stochastic portion of the rubidium oscillator phase; $a(t)$ is a random shock drawn from an ensemble having normal distribution with nonzero mean; B is the backward shift operator, that is $Bx(t) = x(t-1)$; p is the order of the auto-regressive operator; d is the number of integrations (a special case of auto-regression); and q is the order of the moving average operator.

$$(1 - g_1 B - \dots - g_p B^p)(1 - B)^d x(t) = (1 - h_1 B - \dots - h_q B^q) a(t) \quad (1)$$

The auto-regressive portion of the equation has been separated into two components to simplify the analysis. If the parameter p were nonzero, then the solution would contain terms which either vary periodically or are exponentially increasing or decreasing. In the case of atomic clocks, periodic variations are generally attributed to environmental sensitivity and not modelled as a noise process. Thus the ARIMA model for describing an atomic clock in steady state will generally have the parameter $p=0$.

One of the most significant features of rubidium oscillators is that they are characterized by long-term frequency drift which is nearly constant over periods of time for which the clock phase must be predicted but may vary significantly over the life of the device. It has also been found that the noise of a rubidium atomic clock can be characterized as the sum of two independent noise processes - white frequency noise and random walk frequency noise. Thus the second difference of the phase time series of a rubidium oscillator has a nonzero

mean equal to the drift². The ARIMA model corresponding to this situation has two integrations (that is $d=2$). This will result in a phase prediction function which is quadratic in time. The nonzero drift is accounted for by requiring that the random shocks are selected from an ensemble which have a normal distribution with nonzero mean and the problem of estimating that drift reduces to the problem of estimating the mean of the ensemble.

The order of the moving average portion of the ARIMA model can be determined from a consideration of the autocovariance of the noise. The ARIMA model will not adequately describe the rubidium atomic oscillator unless the parameters of the model yield the same autocovariance function of the phase as the observed white and random walk frequency noise processes. The autocovariance function is identically zero for two or more sample intervals between elements (lags). The parameter $q=1$ and the coefficient of the moving average term is given by

$$h_1 = \frac{2\sigma_w^2}{\sigma_{rw}^2 + 2\sigma_w^2 + \sqrt{\sigma_{rw}^4 + 4\sigma_w^2\sigma_{rw}^2}} \quad (2)$$

where σ_w and σ_{rw} correspond to the Allan variances of the white and random walk noise processes respectively at the sampling interval chosen for the ARIMA analysis. Table I shows the variation of the moving average parameter with sampling interval over the range of interest.

TABLE I

Sampling Interval (seconds)	Moving Average Coefficient* h_1
$.17 \times 10^4$.91
1.7×10^4	.38
17×10^4	.01

*The white frequency noise is 3×10^{-11} at 1 s and the random walk frequency noise is 3×10^{-15} at 1 s.

The optimum control equation for the rubidium atomic oscillator phase locked to a reference may now be calculated. For the integrated moving average model with two integrations and $q=1$, the logical form of the control loop is shown in Figure 5. Frequency drift and temperature sensitivity are controlled by feedforward since they are physically deterministic and this method of control introduces the minimum error in the loop. The feedforward control equation provides phase updates based on the initial estimate of phase, the initial estimate of frequency, the current estimate of drift, and the current measurement of temperature. The controllable portion of the phase is the negative of the next predicted clock phase in order to make the error

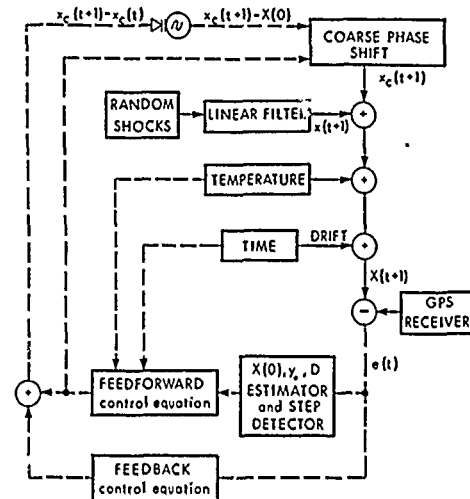


Figure 5.
Feedforward - Feedback control loop configuration.

around the loop zero. Since the drift is removed by the feedforward control equation, the residual random shocks which perturb the loop have zero mean. Thus, the best estimate for the phase of the clock at the next sample is obtained by setting the random shock for that time equal to zero. The resulting control loop equation is given in Equation 3 which we recognize to be the standard form of a second order phaselock loop.

$$X_c(t+1) = - \sum_{t=-\infty}^t [e(t) + (1-h_1) \sum_{t=-\infty}^t e(t)] \quad (3)$$

The required phase corrections are the sum of all the previous errors in the loop plus a sum of all the previous corrections multiplied by a coefficient which varies between zero and unity depending on the white frequency noise level, the random walk frequency noise level, and the sampling time of the loop. Loop performance was evaluated as a function of sampling time using a computer algorithm to simulate the noise in the rubidium oscillator³. The results are plotted in Figure 6 for three different sampling times.

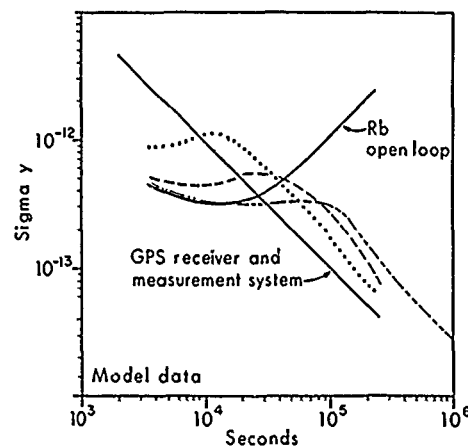


Figure 6.
Simulated Performance of disciplined oscillator for attack times of 4 hours (dotted), 10 hours (dashed), and 24 hours (broken).

The approach which has been used to derive the mixed feedforward feedback control loop for the Rb-GPS disciplined oscillator results in a particularly simple method for loop operation when the time reference is not available. This situation may arise from extended intervals between satellite visibility or GPS receiver failure. Whenever the reference is unavailable, the loop error is set equal to zero and all of the data processing is performed as usual. The integrator in the feedback control equation maintains a constant value while the feedforward control equation continues to update the phase of the oscillator for the previously calculated linear frequency drift and the temperature sensitivity. The optimum estimate of the frequency drift of the rubidium oscillator is simply the mean value of the second difference of the uncorrected phase differences.

IV. Conclusions

A general approach has been developed to control the time of an atomic clock using an external reference and performance data has been presented for a prototype Rb-GPS disciplined oscillator. The technique is quite general and may easily be extended to other oscillators such as quartz crystal oscillators or cesium and hydrogen atomic frequency standards. Similarly, other references such as Loran C or Omega may be used. A designer might expect that a second order phaselock loop will in general be sufficient for any combination of these oscillators and references. However, it is extremely valuable to determine the optimum parameters for the first and second order feedback terms theoretically rather than empirically. Considerable time will be saved by minimizing the need for long term experiments with atomic clocks.

References

1. Box, G. E. P. and Jenkins, GM (1970). "Time Series Analysis Forecasting and Control." Holden Day, San Francisco.
2. Stein, S. R. (1985). Frequency and Time - Their Measurement and Characterization. In "Precision Frequency Control" Vol. II (E. A. Gerber and A. Ballato, eds.) pp. 191-232. Academic Press, New York.
3. Barnes, J. A. (1984). Simulation of Oscillator Noise. Proc. 38th Annual Frequency Control Symposium, pp. 319-326.

SUBMINIATURE RUBIDIUM OSCILLATOR MODEL FRS

Werner Weidemann

Ball Corporation, Efratom Division
18851 Bardeen Avenue, Irvine, CA. 92715 (714)752-2891

ABSTRACT

Efratom has designed and put into production a subminiature rubidium oscillator designated the Model FRS. Emphasis was placed on size (it is presently the smallest rubidium oscillator in the world) as well as on construction suitable for large-scale economic production. This document describes the details of the design concept, key specifications and partial test results.

Our goal was to design an oscillator suitable for most "timebase" applications. This includes all applications where the relative or absolute frequency stability over hours, days, or months is of the utmost importance. Various applications encompass not only frequency counters, but also VLF/Omega and GPS navigation receivers. It is essential that the electronic and mechanical design provide enough protection and rigidity to meet the demands on the oscillator in these applications. The small thermal mass of the oven-controlled components provides fast warm-up, whereas the high thermal resistance reduces the steady-state power requirements.

Significant size reductions were achieved by constructing the microwave cavity out of mu metal and by reducing the size of all submodules/electronics.

INTRODUCTION

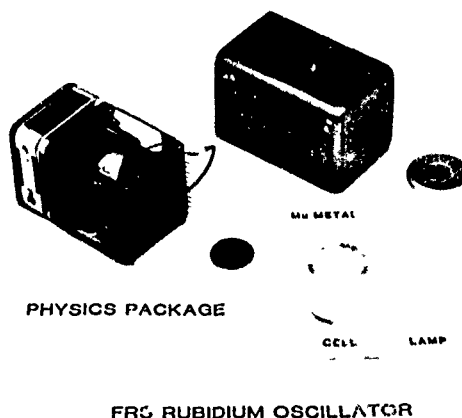
Efratom is the leading manufacturer of rubidium standards. The introduction of its first oscillator, the FRK, in 1972 allowed the use of atomic oscillators in mobile applications. In addition to precise output frequency requirements, these applications place great emphasis on small size, low power, and the ability to perform in a hostile environment. All Efratom rubidium standards utilize an integrated cell containing natural rubidium (Rb) atoms and eliminate the need for a separate filter cell. The problem of potential light shift is solved by an isotopic mixture contained in the rubidium lamp. All of our products also use a TE111 cavity loaded by the glass resonance cell. Both of these features allow a very compact design.

The oscillator must also be price competitive with alternate technologies. To reduce production costs, the number of solder connections between subassemblies must be kept to a minimum. Therefore, the traditional wire harness was eliminated. Only standard off-the-shelf components, including the photocell and mu metal shields, are used.

In order to accommodate a space requirement of 2"x3"x4", a reduction in the size of the physics package (consisting of a lamp section and resonator section) was required. These space requirements also necessitated a reduction in parts count for the electronics section. The following sections describe the design and performance of the FRS in detail.

DESIGN

The Physics Package provides the frequency reference for the crystal oscillator (VCXO). The electronic analogue of the physics package is a high Q ($Q \sim 10^7$) frequency discriminator consisting of a lamp section and resonator section. It is self-contained in that most of the circuitry associated with both sections is located within the physics package, shown in Figure 4.



The Resonator section is comprised of the microwave cavity containing the rubidium resonance cell, the C-field coil and the photocell. The TE111 microwave cavity is made out of silver-plated mu metal, which significantly reduces the size of the package and acts as a magnetic shield. By loading the cavity with the glass cell, the total length was reduced to 20 mm. An aluminum jacket surrounding the cavity acts as heat spreader and a mount for the heater transistors and the thermistor. The cavity thermostat is part of the resonator assembly. The relatively large diameter resonance cell and the low-noise photocell-preamplifier are essential to achieving a signal-to-noise ratio in excess of 70 dB/√Hz.

The Lamp Assembly is shielded from the resonator section and the remainder of the unit. It consists of the lamp-exciter circuit operating at ~80 MHz and the lamp thermostat. As with the

resonator assembly, it is also transistor heated for maximum heater efficiency. An 8 mm spherical lamp (Ball-lamp) is mounted inside the RF coil in such a way as to provide low thermal impedance to the thermostatically-controlled lamp housing. The isotope mixture of rubidium (Rb) in the lamp is adjusted for minimum light shift at operating temperature.

The Crystal Oscillator is built around a 20 MHz 3rd-overtone AT-cut crystal. It features a 1×10^{-5} trim range and can operate over a range of -10°C to $+65^{\circ}\text{C}$ without the need for thermostatic control. An HC-35 crystal housing allows a very compact design. The printed circuit board also contains the 5.3125 MHz frequency synthesizer.

Design emphasis included electrical separation of both circuits through use of proper layout techniques and through use of individual point of use voltage regulators.

The Multiplier triples the 20 MHz signal from the crystal oscillator and phase modulates it at a 127 Hz rate. In order to excite the rubidium ground-state hyperfine transition in the cavity, the 60 MHz driver stage is amplitude modulated at 5.3125 MHz. The hyperfine transition frequency is:

$$\begin{array}{r} 114 \times 60 \text{ MHz} \\ - 5.3125 \text{ MHz} \\ \hline 6834.687500 \text{ MHz} \end{array}$$

A step recovery diode in the microwave cavity generates the 60 MHz comb spectrum as well as the 5.3125 MHz sidebands.

The Servo Assembly is of conventional design with the exception of the sweep circuit and the wide trim range of the crystal oscillator. This trim range is approximately one hundred times wider than the bandwidth of the physics package. A sweep circuit sweeps the crystal through its entire trim range in approximately 15 s. This allows only about 20 ms to observe the 2nd harmonic signal from the physics package which is generated while sweeping through the hyperfine transition frequency.

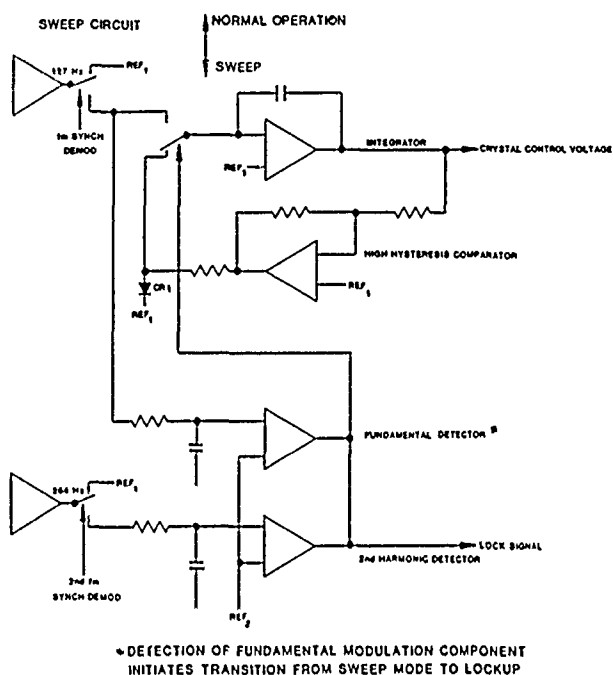
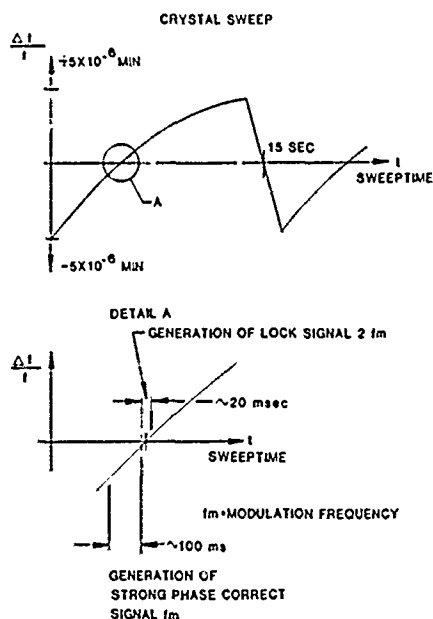


Figure 1. Detail of servo/xtal operation.

In order to allow about 100 ms for the transition from sweep to closed-loop operation, we use the strong fundamental component generated when the crystal frequency approaches f_{RB} . The principle is outlined in Figure 1. Note that the CMOS analog switch controlling sweep/normal operation is activated by the fundamental detector and the 2nd harmonic detector. Diode CR1 and proper selection of the reference voltage, Ref_1 , allows slow up-sweep and fast down-sweep of the crystal frequency.

The Power Supply provides regulated and filtered +17 VDC voltages to all electronic circuits except the heater transistors. The latter are powered directly from the unregulated supply. Key considerations for the power supply are good long-term stability, high ripple rejection, and low noise. The power supply output voltage is increased to 21 VDC until the rubidium lamp ignites. This feature allows operation of the lamp oscillator at lower steady-state power. The 10 MHz output buffer is located on the power supply PCB, in close proximity to the output connector, thereby eliminating the need for a coax cable from the crystal oscillator board to the RF output connector.

FRS Rubidium Oscillator Key Specifications

Output Frequency	10 MHz, .5 Vrms sinusoidal
Long-Term Frequency Drift	5×10^{-10} first year; $2 \times 10^{-10}/\text{yr}$ thereafter
Short-Term Stability	$\sigma_y(\tau) = 5 \times 10^{-11}/(\tau^{-1/2})$ for $\tau \geq 1 \text{ s} < \tau < 100 \text{ s}$
Warm-up	4 min to 1×10^{-9} at 25°C ; 8 min to 1×10^{-9} at -55°C
Input Voltage	22 to 32 Vdc
Input Current (Steady State)	2 A maximum
Operating Temperature	-55°C to $+65^\circ\text{C}$ (Baseplate) ($\Delta f/f < 5 \times 10^{-10}$)
Size	2.0 x 3.0 x 4.0 inch
Weight	1.2 lbs

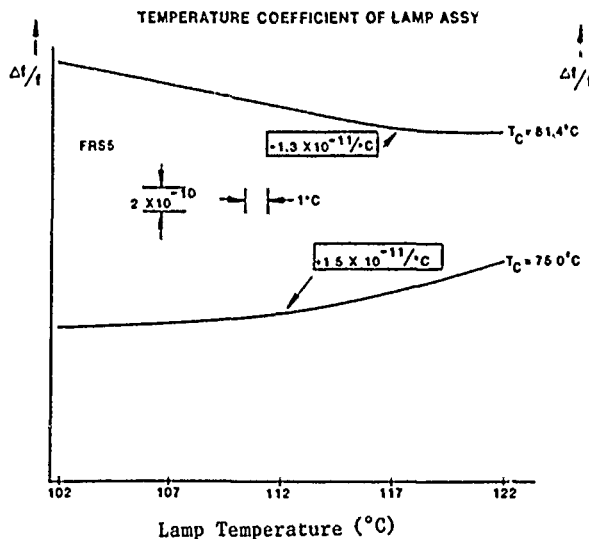
Designed to perform under environments as defined by RTCA DO-160B, including:

- 1g sinusoidal vibration up to 2 kHz
- 1.4 Vrms supply ripple up to 15 kHz
- 95% relative humidity
- 70,000 ft. altitude
- 600 V, 10 μs voltage spikes

PERFORMANCE

All performance parameters, with significant margins, have been verified during pilot production.

The characteristics of lamp temperature vs output frequency and also the dependence of the output frequency vs the resonator temperature are shown in Figure 2.



TEMPERATURE COEFFICIENT OF RESONATOR ASSY

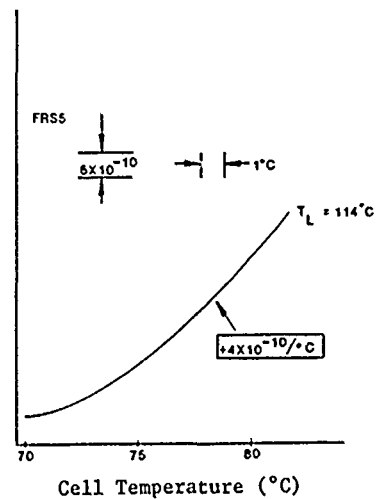
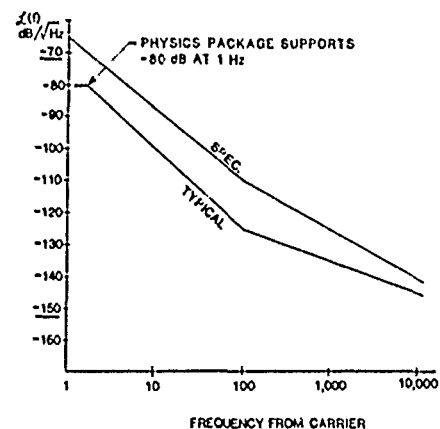


Figure 2. Physics package performance.

Figure 3 shows the typical phase-noise performance of the FRS and the power consumption as a function of ambient temperature.

PHASE NOISE



POWER CONSUMPTION AT 24 VDC (ADD 2 W AT 32 VDC)

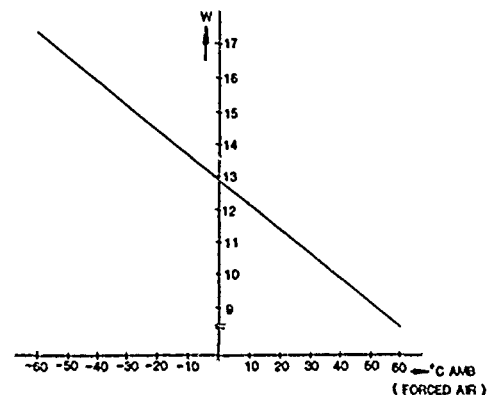


Figure 3. Typical performance.

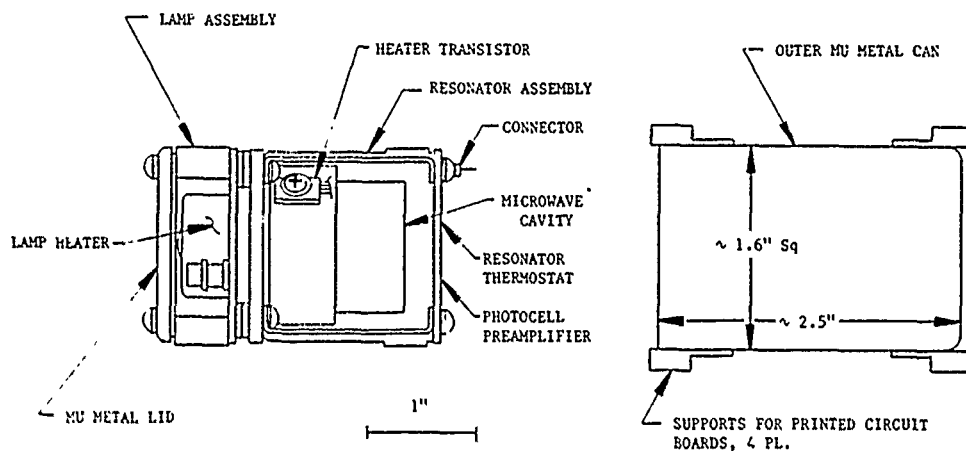


Figure 4. Physics package.

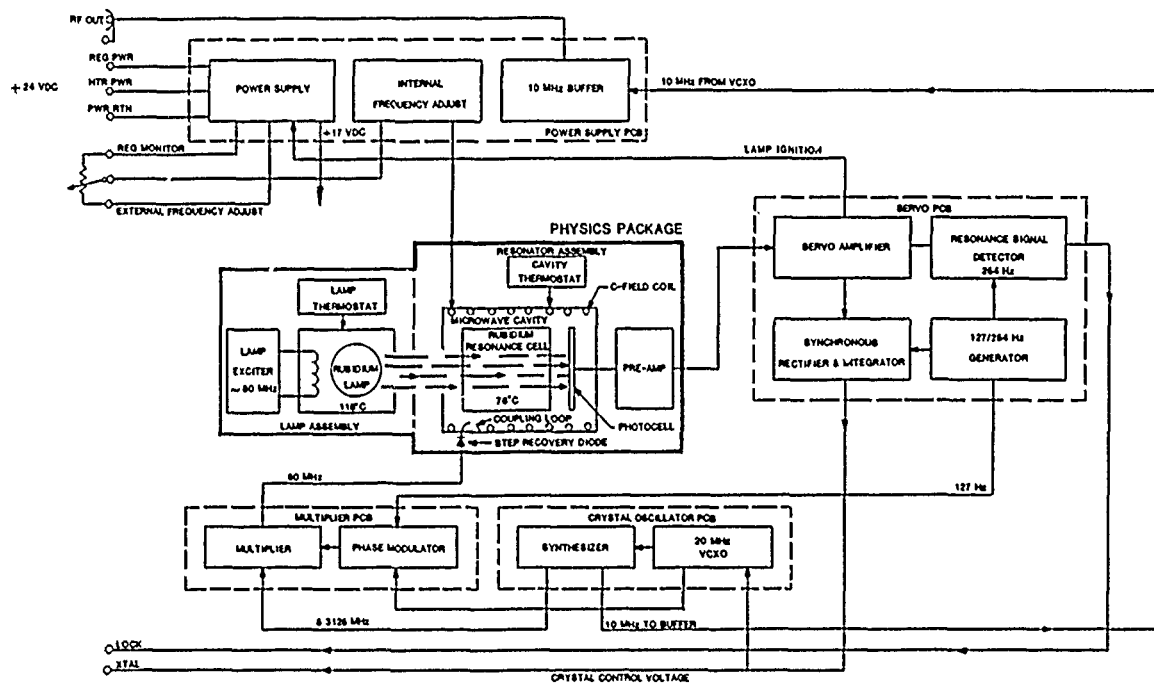


Figure 5. Block diagram.

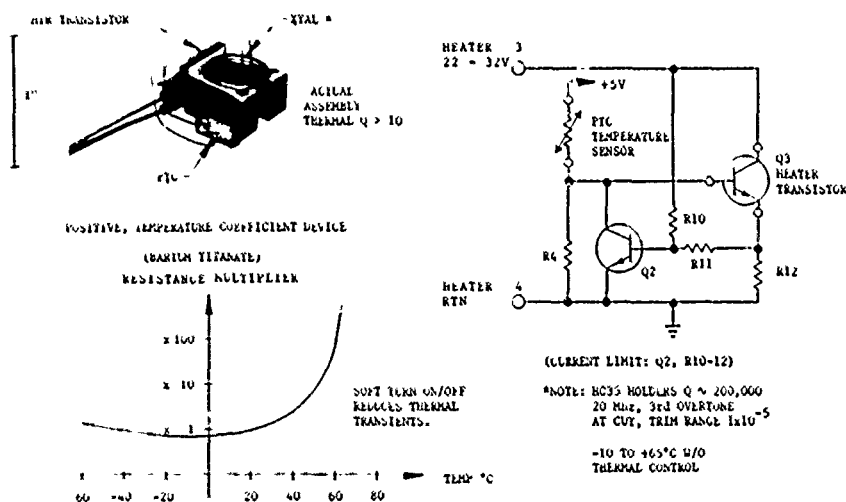


Figure 6. Crystal oscillator thermostat.
(For extended temperature operation only)

AUTHOR INDEX

AUTHOR	PAGE	AUTHOR	PAGE	AUTHOR	PAGE
Adam, J.D.	392	Hardy, W.N.	419	Pucel, R.A.	385
Adler, R.	20, 237	Hatch, E.	145		
Agosta, C.C.	413	Hauden, D.	350	Ramsey, N.F.	6
Aleksic, Z.	340	Hirama, K.	201	Reinhardt, V.	355
Allan, D.W.	394	Holberg, L.W.	428	Richie, S.M.	252
Armington, A.F.	70	Hoskins, M.J.	285	Riley, W.J.	452
Aubry, J.P.	121	Hruska, C.K.	54	Rosenbaum, J.	206
Audoin, C.	410, 432	Hummel, D.C.A.	39		
		Hürimann, M.D.	419	Salvo, Jr., H.L.	206, 392
Balascio, J.F.	70	Hwang, H.B.	32	Schwartzel, J.	101
Ballato, A.	4, 145			Sekimoto, H.	187
Barillet, R.	410	Ihara, T.	187	Shannon, D.L.	127
Benjaminson, A.	344	Iino, T.	224	Sherman, J.H., Jr.	91
Berlinsky, A.J.	419	Iwasaki, F.	39	Shick, D.V.	257, 262
Bernier, L.G.	300	Iwasaki, H.	39	Shimizu, J.	275
Bernstein, M.	18			Shinohara, A.H.	39, 47
Bishop, C.D.	252	Jackson, H.W.	306	Shirley, J.	428
Bottom, V.E.	15	Jani, M.G.	26	Shoji, T.	201
Bower, D.E.	269	Joss, B.	300	Silvera, I.F.	413
Bramble, A.	366	Jumas, J.	101	Smith, W.L.	22
Brandmayr, R.J.	86			Snow, G.S.	230
Brendel, R.	121	Kass, W.J.	230	Stein, S.	465
Brophy, M.J.	285	Kawashima, H.	193	Stokes, R.B.	292
Bustamante, M.	355	Kesperis, J.	366	Suter, J.J.	134
		King, J.C.	16	Suzuki, C.K.	39, 47
Camparo, J.C.	447	Kisenwether, E.C.	373		
Candelier, V.	432	Kiss, J.	47	Takiya, S.	39, 47
Capelle, B.	101	Kohsaka, F.	224	Tamura, F.	325
Cline, R.W.	419	Krishnaswamy, S.V.	206	Tanaka, Y.	201
		Kusters, J.	323	Tang, M.S.H.	152
Dallesasse, J.M.	285			Tellier, C.R.	76
Daly, R.T.	24	Leadon, R.E.	127	Tiersten, H.F.	257, 262
Darces, J.F.	140	Lee, P.C.Y.	152	Toudic, Y.	101
DeMarchi, A.	428, 441	Ling, S.	96	Troxell, W.C.	373
Desmares, P.	237	Lipson, H.G.	63		
Detaint, J.	101	Lukaszek, T.	145	Ueda, T.	224
Dökmeci, M.C.	168				
Doukhan, J.C.	101	MacIntyre, A.	465	Vaccaro, J.R.	452
Dowsett, J.	269	Malocha, D.C.	252	Vanier, J.	432
Doxey, W.L.	9	Marianneau, G.	350	Vasiljevic, D.	340
Driscoll, M.M.	329	Marin, J.J.	26, 32	Vaterkowski, J.L.	76
Drullinger, R.E.	428	Mashimo, Y.	325	Vessot, R.F.C.	413, 422
Dworsky, L.N.	161	Mattison, E.M.	413, 422	Vialle, N.	76
		Maurer, R.H.	134	Viennet, J.	410
Eckart, D.W.	115	McAvoy, B.R.	392	Vig, J.R.	86
EerNisse, E.P.	211, 216	McNab, K.	355		
Elliott, J.H.	292	Mérigoux, H.	140	Walsworth, R.L.	413
Euler, F.K.	295	Miller, M.J.	285	Ward, R.W.	211
Everard, J.K.A.	336	Miura, M.	187	Watanabe, Y.	313
		Mizan, M.	145	Weglein, R.D.	379
Flanagan, T.M.	127			Wei, S.	422
Frueholz, R.P.	447	Nakata, H.	187	Weidemann, W.	470
		Nowick, A.S.	96	Weiss, M.A.	394
Gagnepain, J.J.	121, 350			Wiggins, R.B.	216
Gardiol, F.	300	Ochiai, O.	325	Wilson, T.M.	26, 32
Gerber, E.A.	8	Oliveira, V.A.R.	39, 47	Winkler, G.M.R.	405
Glaze, D.J.	428	Oliver, T.N.	269		
Godfried, H.P.	413	Oomura, Y.	313	Yamamoto, S.	275
Goiffon, A.	101	O'Rourke, D.P.	370	Yamazaki, D.	224
Gould, K.	355			Yen, K.H.	292
Gounji, T.	275	Parker, T.E.	241	Yong, Y.K.	179
Gualtieri, J.G.	115	Pavasovic, A.	340		
		Peterson, J.W.	285	Zarka, A.	101
Hafner, E.	306	Petit, P.	410		
Halliburton, L.E.	26	Philippot, E.	101		

SPECIFICATIONS AND STANDARDS GERMANE TO FREQUENCY CONTROL

INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS

Order through: IEEE Service Center
445 Hoes Lane
Piscataway, NJ 08854
(201) 981-0060

176-1978 Piezoelectricity

177-1966 Piezoelectric Vibrators, Definitions and Methods of Measurements for (ANSI C83.17-1970)

180-1962 Ferroelectric Crystal Terms, Definitions of

319-1971 Piezomagnetic Nomenclature

ELECTRONIC INDUSTRIES ASSOCIATION

Order through: Electronic Industries Association
2001 Eye Street, NW
Washington, DC 20006
(202) 457-4900

(a) Holders and Sockets

RS-192-A, Holder Outlines and Pin Connections for Quartz Crystal Units (Standard Dimensions for Holder types).

RS-367, Dimensional and Electrical Characteristics Defining Receiver Type Sockets (Including crystal sockets).

RS-417, Crystal Outlines (Standard dimensions and pin connections for current quartz crystal units-1974).

(b) Production Tests

RS-186-E, (All Sections), Standard Test Methods for Electronic Component Parts

RS-512, Standard Methods for Measurement of Equivalent Electrical Parameters of Quartz Crystal Units, 1 kHz to 1 GHz, 1985.

(c) Application Information

Components Bulletin No. 6, Guide for the Use of Quartz Crystals for Frequency Control

(d) EIA-477, Cultured Quartz (Apr. 81)

EIA-477-1, Quartz Crystal Test Methods (May 1985)

INTERNATIONAL ELECTROTECHNICAL COMMISSION (IEC)*

Order through: American National Standards Inst. (ANSI)
1430 Broadway
New York, NY 10018

*ANSI can quote prices on specific IEC publications on a day to day basis only. All IEC and ISO standards have been removed from its Standards Catalog. Call ANSI, (212) 354-3300 for prices.

IEC PUBLICATIONS ISSUED:

- 122:- Quartz crystal units for frequency control and selection.
- 122-1 (1976) Part 1. Standard values and test conditions.
- 122-2 (1983) Part 2: Guide to the use of quartz crystal units for frequency control and selection.
- 122-3 (1977) Part 3: Standard outlines and pin connectors.
- 122-3A (1979) First supplement.
- 122-3B (1980) Second supplement.
- 122-3C (1981) Third supplement.
- 283 (1968) Methods for the measurement of frequency and equivalent resistance of unwanted resonances of filter crystal units.
- 302 (1969) Standard definitions and methods of measurement for piezoelectric vibrators operating over the frequency range up to 30 MHz.
- 314 (1970) Temperature control devices for quartz crystal units. Amendment No. 1 (1979)
- 314A (1971) First supplement.
- 368:- Piezoelectric filters.
- 368A (1973) First supplement
Amendment No. 1 (1977)
Amendment No. 2 (1982)
- 368B (1975) Second supplement.
- 368-1 (1982) Part 1: General information, standard values and test conditions.
- 368-3 (1979) Part 3: Standard outlines.
- 368-3A (1981) First supplement.
- 444:- Measurement of quartz-crystal unit parameters by zero phase technique in a π network.
- 444 (1973) Basic method for the measurement of resonance frequency and equivalent series resistance of quartz crystal units by zero phase technique in a π network.
- Note: This publication, where revised, will be issued as a second edition of Publication 444-1.
- 444-2 (1980) Part 2: Phase offset method for measurement of motional capacitance of quartz crystal units.
- 483 (1976) Guide to dynamic measurements of piezoelectric ceramics with high electromechanical coupling.
- 642 (1979) Piezoelectric ceramic resonators and resonator units for frequency control and selection -
Chapter I : Standard values and conditions.
Chapter II: Measuring and test conditions.

- 679:- Quartz crystal controlled oscillators.
- 679-1 (1980) Part 1: General information, test conditions and methods.
- 679-2 (1981) Part 2: Guide to the use of quartz crystal controlled oscillators.
- 689 (1980) Measurements and test methods for 32 kHz quartz crystal units for wrist watches and standard values.
- 758 (1983) Synthetic quartz crystal - Chapter I: Specification for synthetic quartz crystal - Chapter II: Guide to the use of synthetic quartz crystal.

DEPARTMENT OF DEFENSE

Order through: Naval Publication & Form Center
5801 Tabor Avenue
Philadelphia, PA 19120

MIL-C-3098 Crystal Unit, Quartz, Gen Spec for
MIL-H-10056 Holders (End), Crystal, Gen Spec for
MIL-STD-683 Crystal Units, Quartz/Holders, Crystal
MIL-O-55310 Oscillators, Crystal, Gen Spec for
MIL-F-18327 Filters, High Pass, Low Pass, Band Pass
Suppression and Dual Functioning, Gen Spec for
MIL-O-39021 Oven, Crystal, Gen Spec for
MIL-O-55240 Oscillators, Audio Frequency
MIL-F-28734 Frequency Standards, Cesium Beam, Gen Spec
MIL-F-288111 Frequency Standard, Cesium Beam Tube
MIL-C- 24523 (SHIPS), Chronometer Quartz Crystal

PROCEEDINGS

ANNUAL FREQUENCY CONTROL SYMPOSIA

<u>NO.</u>	<u>YEAR</u>	<u>DOCUMENT NUMBER</u>	<u>OBTAIN FROM*</u>	<u>COST</u>
10	1956	AD298322	NTIS	\$41.50
11	1957	AD298323	"	44.50
12	1958	AD298324	"	46.00
13	1959	AD298325	"	49.00
14	1960	AD246500	"	32.50
15	1961	AD265455	"	28.00
16	1962	PB162343	"	35.50
17	1963	AD423381	"	43.00
18	1964	AD450341	"	43.00
19	1965	AD471229	"	47 50
20	1966	AD800523	"	47.50
21	1967	AD659792	"	41.50
22	1968	AD844911	"	44.50
23	1969	AD746209	"	25.00
24	1970	AD764210	"	28.00
25	1971	AD746211	"	28.00
26	1972	AD771043	"	26.50
27	1973	AD771042	"	34.00
28	1974	ADA011113	"	31.00
29	1975	ADA017466	"	34.00
30	1976	ADA046089	"	40.00
31	1977	ADA088221	"	44.50
32	1978		EIA	20.00
33	1979		"	20.00
34	1980		"	20.00
35	1981		"	20.00
36	1982	ADA130811	NTIS	41.50
37	1983	83CH1957-0	IEEE	59.00
38	1984	84CH2062-8	"	59.00
39	1985	85CH2186-5	"	59.00
40	1986	86CH2330-9	"	59.00

* NTIS - National Technical Information Service
Sills Building
5285 Port Royal Road
Springfield, VA 22161

* EIA - Annual Frequency Control Symposium
c/o Electronic Industries Association
2001 Eye Street
Washington, DC 20006

*IEEE - Institute of Electrical & Electronics Engineers
445 Hoes Lane
Piscataway, NJ 08854

Remittance must be enclosed with all orders. Prices are subject to change without prior notice.

A subject and author index for the Proceedings of the 10th through the 38th Symposia appears as a supplement to the 38th Proceedings volume.